University of Rhode Island
DigitalCommons@URI

Open Access Dissertations

2016

Development and Validation of Immunoinformatics Tools for Swine Vaccine Development

Andres H. Guiérrez Núñez
University of Rhode Island, ahgutierrezn@uri.edu

Follow this and additional works at: https://digitalcommons.uri.edu/oa_diss

Recommended Citation

Guiérrez Núñez, Andres H., "Development and Validation of Immunoinformatics Tools for Swine Vaccine Development" (2016). *Open Access Dissertations.* Paper 526.
https://digitalcommons.uri.edu/oa_diss/526

This Dissertation is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons@etal.uri.edu.

DEVELOPMENT AND VALIDATION OF IMMUNOINFORMATICS TOOLS FOR SWINE VACCINE DEVELOPMENT

BY

ANDRES H. GUTIÉRREZ NÚÑEZ

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

CELL AND MOLECULAR BIOLOGY

UNIVERSITY OF RHODE ISLAND

2016

DOCTOR OF PHILOSOPHY DISSERTATION

OF

ANDRES H. GUTIÉRREZ NÚÑEZ

APPROVED:

Dissertation Committee:

Major Professor — Anne S. De Groot

Co-Major Professor — Leonard Moise

Alan Rothman

Marta Gomez-Chiarri

Chris Bailey-Kellogg

Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2016

ABSTRACT

Immunoinformatics tools have multiple applications in human immunology research. One of their most prominent applications is the prediction of T cell epitopes to accelerate vaccine development. T cell epitopes are short peptides derived from pathogens that are recognized by the cells of the immune system (T cells) when presented on the surface of cells bound to a major histocompatibility complex (MHC) molecule, which results in a specific immune response. For their role as key drivers of the immune response, numerous algorithms have been developed to predict binding of peptides to MHC molecules; however, comparable tools are limited for other species. The goal of this thesis is to develop immunoinformatics tools for swine to aid in the design of vaccines for pathogens affecting the pork industry. One of the main reasons for the limited development of T cell epitope mapping tools for swine is the lack of data required to train and test the predictive algorithms. Through this research, we have developed PigMatrix, a tool for prediction of peptide binding to swine MHC molecules. In an initial analysis, PigMatrix predictive performance was favorable, in particular because its development did not require training data. Using PigMatrix, we have identified immunogenic peptides conserved in seven different strains of influenza A virus (IAV), a highly diverse virus that has a significant impact not only on swine, but also for humans. Protective potential of IAV vaccines is commonly predicted using genetic data and antibody cross-reactivity properties of the hemagglutinin (HA) surface protein, the most variable antigen and primary target of the antibody immune response. However, protection has been reported in the absence of cross-reactive antibodies to HA. To explore the role of T cell epitopes in vaccine protection, we have developed a method (EpiCC) to compare T cell epitope content between proteins. We found that the relationship of predicted T cell epitopes between HA sequences of a swine IAV inactivated vaccine and challenge strains was associated with protection, providing evidence that T cells contribute to vaccine efficacy. This approach may complement current methods for selection of influenza vaccines against novel viruses and influenza strains for vaccine development. Taken together, these findings demonstrate the potential of immunoinformatics tools for the development and evaluation of swine vaccines and will allow for further research to improve the tools and apply them to design novel vaccine candidates.

ACKNOWLEDGMENTS

I would like to thank my dissertation advisors, Dr. Annie De Groot and Dr. Lenny Moise for the opportunity to join the Institute for Immunology and Informatics, for providing an example of tireless dedication to science but not to the status quo, and for offering challenges that taught me to think creatively and independently.

Thank you to the members of my thesis committee for their time, feedback, and advice while serving on my thesis committee: Dr. Joan Peckham (Department of Computer Sciences and Statistics), Dr. Marta Gomez-Chiarri (Department of Fisheries, Animal and Veterinary Sciences), Dr. Alan Rothman (Department of Cell and Molecular Biology), and Dr. Chris Bailey-Kellogg (Department of Computer Science, Dartmouth College). In addition, I would like to thank the chair of my committee, Dr. Rachel Schwartz (Department of Biological Sciences).

To all the members of the Institute for Immunology and Informatics, thank you for putting up with me for all these years, and for always being available to help and support me through this process.

Thank you to Bill Martin and the EpiVax team for patiently teaching me about immunoinformatics, for being a valuable sounding board for new ideas and new approaches to problems, and for the time they have invested in reading, reviewing and deploying my work.

I must thank Dr. Crystal Loving for her dedication to our work together, and for the experience of research in her laboratory at the USDA, where I learned so much about immunology and working firsthand with pigs. To Zahra Olson, thank you for explaining the details of experiments and protocols with the patience only a good friend can provide. I gratefully acknowledge the sacrifice of the piglets without which our research would not be possible.

Thank you to Danielle Aguirre for sharing the ups and downs of being a graduate student at the Institute for Immunology and Informatics, and in the process becoming a great friend who is always on my team.

A mi familia, gracias por estar siempre a mi lado ignorando lo que dice la geografía. Por ser mi motivación para levantarme después de cada caída y esforzarme para ser mejor cada día. Gracias por ser ejemplo de trabajo incansable y por enseñarme que alcanzar mis metas requiere esfuerzo y sacrificio. Gracias por apoyar mis decisiones y dejarme volar lejos. Gracias a mi abuelito Manuel por cuidarme y creer en mí siempre. Frances, no lo habría poder haber hecho sin ti. Gracias por ser mi fuente constante de alegría y mi apoyo incondicional.

PREFACE

The following dissertation has been prepared in manuscript format according to the guidelines of the Graduate School of the University of Rhode Island. The dissertation includes five chapters: Chapter 1, Introduction. Chapter 2, "Development and validation of an epitope prediction tool for swine (PigMatrix) based on the pocket profile method", was published in *BMC Bioinformatics*. Chapter 3, "In vivo validation of predicted and conserved T cell epitopes in a swine Influenza model" was published in *PLoS ONE*. Chapter 4, "T cell epitope Content Comparison (EpiCC) of swine H1 influenza A virus hemagglutinin" was prepared for submission to *Influenza and Other Respiratory Viruses*. Chapter 5, Conclusions.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

APPENDIX: Supplemental information for Chapter 4

CHAPTER 1

Introduction

Immunoinformatics, also known as computational immunology, is a branch of bioinformatics that applies computational methods for both basic and translational immunology and vaccine research. One of the most common applications of immunoinformatics is the prediction of T cell epitopes. T cell epitopes are short linear peptides derived from pathogen proteins (or self proteins) that are recognizable by cells of the immune system (T lymphocytes or T cells) and induce a specific adaptive T cell-mediated immune response (CMI) [1,2]. They are recognized by specialized antigen receptors on T cells, T cell receptors (TCRs); however, TCRs only recognize peptides if they are bound to major histocompatibility complex (MHC) molecules and presented on the surface of cells [2,3]. Due to the key role of T cells in the adaptive immune response against pathogens, numerous immunoinformatics tools have been developed to predict human T cell epitopes [4]. However, similar tools are limited for animals such as swine, cattle, and other important livestock.

MHC molecules

There are two classes of MHC molecules that interact with peptides; MHC class I and MHC class II [3]. These molecules are involved in different antigen presentation pathways and play different roles in the immune response to pathogens [2,3].

MHC Class I

MHC class I molecules are expressed by all nucleated cells and generally present peptides derived from intracellular proteins of self and non-self origin [3]. The proteins are initially degraded into peptides by the proteasome and translocated into the endoplasmic reticulum (ER) by the transporter associated with antigen processing (TAP) [3,5]. Suitable peptides bind semi-folded MHC class I molecules anchored in the ER, and induce further folding facilitated by the peptide-loading complex [3,5,6]. Peptide-MHC complexes are then transported to the cell surface where they are presented to CD8 T cells circulating in the lymphatic system [3]. Upon TCR recognition of foreign peptides and binding of the T cell surface co-receptor CD8 to a second constant site on MHC class I, mature naïve CD8 T cells become activated and differentiate into cytotoxic effector CD8 T cells (cytotoxic T lymphocytes or CTL) [7]. CTL leave

the lymphoid tissue in high numbers and migrate to infected tissue [8]. The major roles of CTL are to (1) eliminate infected target cells by inducing apoptosis through release of cytotoxic proteins (perforins, granzymes) or expressing Fas ligand that binds to Fas on target cells, and (2) secrete cytokines (IFNγ, TNF-α and TNF-β), which contribute to host defense [2]. Once the infection is cleared, most of the CTL die, but a minor subset survive and become memory CD8 T cells. These cells are capable of responding rapidly to any subsequent similar infection [7].

MHC Class II

MHC class II molecules are expressed on professional antigen presenting cells (APC) and generally present peptides derived from extracellular proteins [2,3,9]. The proteins are first taken into the cell by phagocytosis or endocytosis, after which endocytic vesicles become acidified, fuse with lysosomes and then antigens are degraded by proteases (e.g. cathepsins) [3,9]. MHC class II molecules are assembled in the ER with the invariant chain (Ii) in the binding groove. MHC-Ii complexes are transported to endocytic vesicles (MIIC) where the Ii undergoes proteolysis, leaving in the cleft the MHC class II-associated invariant chain peptide (CLIP). In the MIIC, antigenic fragments generated by resident proteases may bind MHC class II. CLIP has to be exchanged with peptides of higher affinity. CLIP removal is facilitated by HLA-DM. Peptide-MHC complexes are transported to the cell surface for presentation to CD4 T cells in the lymphoid system [3,9,10]. TCR recognition of pathogen-derived peptides bound to MHC, binding of the co-receptor CD4 and an additional co-stimulatory signal (B7:CD28 interaction), lead to the activation of naïve CD4 T cells, which then differentiate into distinct CD4 T helper effector subsets that produce diverse cytokines and differentially drive the immune response. The currently known types of T helper (Th) cells are: Th1 cells, Th17 cells, Th2 cells, T follicular helper cells (Tfh), and regulatory T cells (Treg) [7]. Th1 cells secrete IFNγ, which activates macrophages to respond against intracellular viral and bacterial infections. Th17 cells secrete IL-17 to activate and recruit neutrophils to fight extracellular bacterial and fungal infections. Th2 cells secrete IL-4 that acts on mast cells, basophils, and eosinophils to respond to parasite infections. Tfh cells activate naïve B cells to differentiate into antibody-producing cells. Treg suppress other effector CD4 and CD8 T cells. Like CTL, most antigen-specific Th die after resolution of the infection, but a small percentage of the CD4 Th cells convert to memory cells [7].

The antigen presentation pathways described above are simplified and exceptions have been reported [11]. There is evidence that demonstrates that MHC class I and II pathways are not always separated and that peptides derived from extracellular antigens can end up being presented by MHC class I molecules (i.e. cross-presentation) [12].

There are important structural differences between the MHC class I and II proteins (Fig. 1) that determine the specificity of peptide binding. Class I molecules consist of a transmembrane (α) chain and a β2-microglobulin chain linked non-covalently. The α chain has three structural domains (α1, α2 and α3); the α1 and α2 domains form the peptide-binding groove that is closed at the ends, which limits the length of peptides able to bind [13]. Ninemer peptides (9-mers) are generally preferred for binding, but shorter and longer peptides have been reported [3]. Class II molecules consist of two transmembrane chains (α and β). The α1 and β1 domains form an open ended binding groove. Hence, longer peptides, generally 13-25 amino acids in length, bind MHC class II [13]. Similarly to class I, only nine amino acids fit into the groove; the remaining residues on both ends play a role in stabilizing the interaction [14,15].

MHC genes

Human MHC genes

The genes that encode MHC class I and II proteins are highly polymorphic. In humans, the α chains of classical class I molecules are encoded by genes in the human leukocyte antigen (HLA)-A, HLA-B and HLA-C loci. Class II chains are encoded by genes in the HLA-DR, HLA-DQ and HLA-DP loci. HLA-DRβ chain can be encoded by four loci (HLA-DRB1 and DRB3-5); however, only HLA-DRB1 is expressed in all individuals. The HLA-DRα chain, which exhibits limited polymorphism, is expressed from a single locus (HLA-DRA).

Thousands of HLA-A, HLA-B and HLA-C alleles have been reported as well as more than a thousand HLA-DRB1 alleles [16]. Most of the differences between these alleles can be attributed to sequence polymorphisms in the binding groove that determine which peptides will be able to bind to a particular MHC allele. Therefore, different alleles generally have different binding preferences, although several groups of alleles may have similar preferences [17–19]. Despite their diversity, some

alleles are more frequently expressed. For example, HLA-A*0201 is expressed by 47% of the European population based on an allele frequency database [20].

Swine MHC genes

Pigs also express MHC, called swine leukocyte antigen (SLA). Three loci encode the classical SLA class I (SLA-1, SLA-2, SLA-3) [21]. The expression level of the SLA-1 gene is the highest whereas SLA-3 is the lowest [22]. Several loci encode genes for expression of SLA class II proteins including α and β chain genes for SLA-DR and SLA-DQ. Pigs do not express DP proteins. As is true for human HLA-DRB1, SLA-DRB1 is commonly expressed, and like HLA-DRA, the SLA-DRA locus is highly conserved [21]. Current methods for the characterization of SLA alleles rely on DNA-based low- and high-resolution PCR sequence specific primers (PCR-SSP). This method uses allele-specific primers that cover polymorphic sites unique to a given allele [23]. While high-resolution PCR-SSP identifies specific SLA alleles, low-resolution SLA-typing differentiates alleles by groups that share similar sequence motifs [24,25]. SLA molecules are also highly polymorphic. As of October 2016, 116 SLA class I alleles (including SLA-1, SLA-2 and SLA-3 loci) and 82 SLA-DRB1 allele sequences have been deposited in the Immune Polymorphism Database (www.ebi.ac.uk/ipd/mhc/group/SLA).

Experimental identification of swine T cell epitopes

Several experimental approaches have been developed to evaluate HLA allele-specific binding preferences and map human T cell epitopes including intracellular cytokine staining, T cell proliferation, enzyme-linked immunosorbent spot (ELISpot), and HLA binding assays [26]. In vitro restimulation of T cells with overlapping peptides is the method most commonly used to identify swine immunogenic T cell peptides [27–36]. In this assay, peripheral blood mononuclear cells (PBMCs; consisting of T cells, B cells, dendritic cells, and natural killer cells) isolated from donor blood, exposed to a particular antigen by natural infection or vaccination, are stimulated with peptides. Peptides of 15-20 residues, overlapping by at least 8 amino acids to cover all potential epitope 9-mer cores of an antigen, are used for stimulation. If the immune system has encountered a given epitope previously, a specific recall T cell response can be detected by the amount of released cytokines. IFNγ ELISpot assay is widely used in pigs to measure the frequency of antigen-specific cytokine-secreting cells [32–36].

Although the ELISpot assay is a very sensitive method to measure epitope-specific immunogenic responses, its application to identification of T cell epitopes has some limitations. First, under certain conditions, T cell responses may be limited to a subset of dominant epitopes. For example, responses to subdominant epitopes may only be detected if the dominant epitopes are absent from the antigen during exposure or vaccination [37]. Second, T cells that recognize peptides bound to MHC might induce cytokines different from the one being tested [26,38]. Third, ELISpot assays do not provide information about the phenotype of cells producing IFNγ [26,38].

SLA binding assays have also been applied to map swine class I T cell epitopes [39–42]. These assays quantify peptide-MHC affinity, which makes this approach appropriate for identification of peptide binders and non-binders. Competition assays using a labeled reference peptide bound to an MHC molecule and increasing concentrations of the query peptide are available for several HLA class I and II alleles, but they have been developed for only three SLA class I alleles [39–42].

T cell epitope prediction tools

There is a large and ever-increasing body of experimental data on MHC-peptide binding measurements that has been accumulated over the past few decades, which has enabled the development of highly accurate T cell epitope prediction algorithms for humans [43]. Epitope-prediction methods use these large sets of quantitative binding data to first train and then test HLA allele-specific models. Support vector machines (SVMs), hidden Markov models (HMMs) and artificial neural networks (ANNs) have been applied to develop predictions for characterized HLA alleles. For the purposes of predicting T cell epitopes for broad populations of subjects, HLA alleles are commonly clustered into "HLA supertypes", based on their binding specificities, reducing the complexity generated by the polymorphism of HLA genes [17–19].

For uncharacterized MHC molecules with limited or nonexistent experimental data, methods that infer binding preferences using information about multiple well-studied MHC alleles have been developed. A common assumption of these "pan-specific" methods is that MHC molecules with similar contact residues in their binding grooves have similar binding preferences [44].

The crystallographic structure of the HLA molecules revealed that the peptide-binding groove contains a number of pockets and that polymorphic residues in the HLA sequence are often involved in forming these pockets [13,45]. Consequently, the residues in the pocket define allele-specific binding preferences for particular amino acid side chains of the antigenic peptides [46]. Thus, for each MHC molecule, the profile of a given binding pocket can be defined by its residues and binding preferences [47]. Sturniolo et al. demonstrated that each "pocket profile" was nearly independent of the remaining HLA-DR binding groove [47]. The authors also showed that an MHC molecule could be defined in terms of its individual pocket profiles as a quantitative matrix of binding preferences. Therefore, once a pocket profile is determined experimentally, it can be shared with other HLA-DR molecules that have identical pocket residues [47].

The pocket profile method has been applied to develop different pan-specific algorithms for T cell epitope prediction including TEPITOPE [47], TEPITOPEpan [48], PickPocket [49], and EpiMatrix [50]. Previous studies showed that the predictive performance of pan-specific methods for novel HLA alleles depends on the similarity of the pocket residues; the performance normally decreases as the similarity decreases [49]. For HLA alleles with limited quantitative data, algorithms based on the pocket profile method had better or comparable performance than methods that require a large amount of data for training (e.g. ANNs) [48,49].

Thesis motivation and outline

The most frequent application of T cell epitope prediction tools is for identification of immunogenic peptides to design and develop vaccines against a wide variety of infectious pathogens [51–54]. They have also been applied for development of therapeutic vaccines [54–56], diagnostic tests [57], deimmunization of biological drugs [58,59], research related to transplantation [60], allergy [61], and autoimmunity [62]. Although epitope-based vaccines have yet to be licensed, several clinical trials involving epitope-based strategies are underway (clinicaltrials.gov) [63].

Despite their multiple applications and ability to reduce time, effort and resources required to identify T cell epitopes [64], prediction algorithms are scarcely available for non-human species. This is mainly due to the limited experimental data available to develop predictors and may also be due to lack

of interest in epitope-driven vaccines for livestock animals. However, epitope-prediction tools for swine could have at least as many applications as for humans and would potentially accelerate porcine immunology research. One online tool is available for prediction of SLA class I epitopes [65]. This algorithm was trained and evaluated for prediction of three SLA class I alleles [40,65]. Prediction tools are not available for SLA class II alleles.

To address the paucity of epitope prediction tools, Chapter 2 of this thesis describes the application of the pocket profile method to SLA alleles using binding preferences previously defined for HLA molecules to develop "PigMatrix". This approach was developed to overcome the lack of quantitative data required to construct swine T cell epitope predictors. The pocket residues of SLA molecules were defined based on swine and human MHC crystal structures. Different scoring systems were tested to determine similarities between SLA and HLA pocket sequences and to identify the best human match for the purpose of inferring binding preferences of each SLA pocket in the binding groove. Using this information, matrices were built and their predictive performed tested retrospectively using datasets of SLA-restricted epitopes that were available for one SLA class II allele and three SLA class I alleles. For SLA class I alleles, the predictive power of PigMatrix was comparable to or higher than two other available methods. PigMatrix is the first tool available for prediction of SLA class II epitopes. These results demonstrated the potential of the pocket profile method and PigMatrix to develop predictions for SLA alleles without any training step involved, leveraging HLA binding preferences in order to circumvent training on published ligands as is generally required for quality predictive models.

In Chapter 3, additional predictive matrices were developed for SLA class I and II alleles that have been reported to be prevalent in the U.S. swine population [24,25], and applied to identify potentially immunogenic peptides conserved in seven swine influenza A viruses. The immunogenicity (i.e. ability to induce an immune response) of 48 predicted T cell epitopes was determined by measuring IFNγ recall responses using PBMCs from pigs immunized with prototype DNA vaccines designed to deliver the peptides. Positive responses that were observed upon restimulation with pooled peptides as well as eleven individual peptides demonstrated that PigMatrix prediction identified SLA class I and II epitopes. However, it was shown after sacrifice of the animals that there was a mismatch

between the SLA alleles expressed by the animals in the study and the set of alleles utilized for the selection of the peptides. Therefore, cohort-specific SLA epitope predictors were developed and used to re-evaluate binding likelihood of the peptides. Retrospectively, cohort-specific PigMatrix analysis was able to predict non-immunogenic peptides efficiently.

Influenza A virus (IAV) is an attractive model pathogen to evaluate swine T cell epitope predictions for several reasons. Importantly, IAV causes a highly contagious disease in pigs that has a significant impact to the swine industry mainly due to its high morbidity and reduction in the growth rate of infected pigs, complicating the management of commercial processing [66,67]. Furthermore, IAV is a zoonotic pathogen that can be transmitted between pigs and people, which is a matter of public health concern [68]. Swine-origin IAVs can adapt to the human host and spread between humans resulting in a pandemic, as was demonstrated during the 2009 influenza A (H1N1) pandemic [68,69]. Due the impact of IAV infections in pigs and humans, effective vaccines are highly desirable.

Development of IAV vaccines is a significant challenge due to the variability of the virus. IAVs are enveloped viruses with negative-sense, single stranded, segmented RNA genomes. The eight RNA segments of IAV encode two nonstructural proteins (NS1 and NS2), the nucleoprotein (NP), three RNA polymerase proteins (PA, PB1, and PB2), two polypeptides synthesized from the PB1 mRNA (PB1-F2 and PB1-N40), the novel PA-X, two matrix proteins (M1 and M2), and two surface glycoproteins, hemagglutinin (HA), neuraminidase (NA); these two surface proteins determine different IAV subtypes [69]. Three different subtypes of IAV (H1N1, H1N2 and H3N2) co-circulate and are predominant in the North American swine population [70]. Reassortment of RNA segments from different viral strains infecting the same cell (antigenic shift) generates novel IAV strains [70]. Pigs are considered a “mixing vessel” that produces new IAV strains; they act as an intermediate host in cross-species transmission of IAV because their respiratory epithelium expresses HA receptors that bind to both avian and human IAV strains [71]. In fact, most swine influenza viruses are reassortants composed of mixtures of human, avian and swine virus genes [70]. Accumulation of mutations in HA and NA (antigenic drift), also contributes to the diversity of swine IAV [70].

Humoral immune response (antibodies) and CMI (comprised of responses from CTL and Th cells) participate in the adaptive immune response against IAV [72]. Neutralizing antibodies that target

HA block the attachment of this protein to sialic receptors in the host cells preventing infection [73]. Vaccine protection, mediated by anti-HA antibodies, against a different viral strain depends on the similarity between the HA proteins of the two viruses. Therefore, variability of HA can reduce the protective effects of pre-existing antibodies [73]. On the other hand, CMI can be broadly cross-reactive and protective against variety of IAV subtypes [74–78]. CMI does not prevent infection, but it does contribute to viral clearance of IAV as well as to reduced clinical signs and viral shedding in humans and mice [79,80]. Cell-mediated immune responses have also been observed in the lungs of pigs infected naturally [81], vaccinated with whole inactivated virus vaccine or live-attenuated vaccines [82–85]. Reduction of morbidity is an important objective for swine IAV vaccines to prevent lack of appetite and loss of bodyweight.

The genetic relationships and cross-reactive properties of antibodies to the HA proteins of the vaccine viruses are widely used to predict cross-protection. The hemagglutination inhibition (HI) assay measures the concentration of serum anti-HA antibodies against a given IAV strain and is commonly used for predicting protective immunity mediated by neutralizing antibodies to HA [86,87]. However, there are studies in pigs that show protection (reduced lung lesions, reduced viral titers in lungs and/or nasal swabs) conferred by swine vaccines even in the absence of cross-reactive antibodies [82,87–91]. In addition, the genetic lineage of the virus does not always predict antigenic phenotype. Therefore, new means of predicting the protective efficacy of swine IAV vaccines are needed.

To determine the potential role of T cell epitope-driven CMI in protection, in Chapter 4, a method for comparison of predicted T cell epitope content (EpiCC) was developed. This method was applied to measure the relatedness, at the T cell epitope level, between the HA from swine IAV strains representing the major H1 clusters circulating in the North American swine population and those of H1 viruses in a commercial vaccine. Using experimental data from previous efficacy studies testing one of the H1 viruses in the commercial vaccine against different challenge viruses [82,87–91], a certain level of T cell epitope relatedness associated with protection was identified. The results showed that the T cell epitope content relationship deduced from vaccine efficacy studies provides evidence that T cells contribute to vaccine efficacy. More vaccine efficacy data using diverse vaccine and challenge strains will be required to further validate and improve the predictive potential of this approach.

References

1. Pappalardo F, Flower D, Russo G, Pennisi M, Motta S. Computational modelling approaches to vaccinology. Pharmacol. Res. 2015;92:40–5.

2. Parham P. The Immune System. Fourth Edition. Garland Science; 2014.

3. Neefjes J, Jongsma MLM, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat. Rev. Immunol. 2011;11:823–36.

4. Patronov A, Doytchinova I. T-cell epitope vaccine design by immunoinformatics. Open Biol. 2013;3:120139.

5. Mester G, Hoffmann V, Stevanović S. Insights into MHC class I antigen processing gained from large-scale analysis of class I ligands. Cell. Mol. Life Sci. CMLS. 2011;68:1521–32.

6. Purcell AW. The peptide-loading complex and ligand selection during the assembly of HLA class I molecules. Mol. Immunol. 2000;37:483–92.

7. Kaech SM, Wherry EJ, Ahmed R. Effector and memory T-cell differentiation: implications for vaccine development. Nat. Rev. Immunol. 2002;2:251–62.

8. Masopust D, Schenkel JM. The integration of T cell migration, differentiation and function. Nat. Rev. Immunol. 2013;13:309–20.

9. van den Hoorn T, Paul P, Jongsma MLM, Neefjes J. Routes to manipulate MHC class II antigen presentation. Curr. Opin. Immunol. 2011;23:88–95.

10. Roche PA, Furuta K. The ins and outs of MHC class II-mediated antigen processing and presentation. Nat. Rev. Immunol. 2015;15:203–16.

11. Stern LJ, Santambrogio L. The melting pot of the MHC II peptidome. Curr. Opin. Immunol. 2016;40:70–7.

12. Joffre OP, Segura E, Savina A, Amigorena S. Cross-presentation by dendritic cells. Nat. Rev. Immunol. 2012;12:557–69.

13. Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, et al. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. Nature. 1994;368:215–21.

14. Godkin AJ, Smith KJ, Willis A, Tejada-Simon MV, Zhang J, Elliott T, et al. Naturally processed HLA class II peptides reveal highly conserved immunogenic flanking region sequence preferences that reflect antigen processing rather than peptide-MHC interactions. J. Immunol. Baltim. Md 1950. 2001;166:6720–7.

15. Lovitch SB, Pu Z, Unanue ER. Amino-terminal flanking residues determine the conformation of a peptide-class II MHC complex. J. Immunol. Baltim. Md 1950. 2006;176:2958–68.

16. Robinson J, Halliwell JA, Marsh SGE. IMGT/HLA and the Immuno Polymorphism Database. Methods Mol. Biol. Clifton NJ. 2014;1184:109–21.

17. Southwood S, Sidney J, Kondo A, del Guercio MF, Appella E, Hoffman S, et al. Several common HLA-DR types share largely overlapping peptide binding repertoires. J. Immunol. 1998;160:3363–73.

18. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. BMC Immunol. 2008;9:1.

19. Wang M, Claesson MH. Classification of human leukocyte antigen (HLA) supertypes. Methods Mol. Biol. Clifton NJ. 2014;1184:309–17.

20. González-Galarza FF, Takeshita LYC, Santos EJM, Kempson F, Maia MHT, da Silva ALS, et al. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. Nucleic Acids Res. 2015;43:D784-788.

21. Lunney JK, Ho C-S, Wysocki M, Smith DM. Molecular genetics of the swine major histocompatibility complex, the SLA complex. Dev. Comp. Immunol. 2009;33:362–74.

22. Smith DM, Lunney JK, Martens GW, Ando A, Lee J-H, Ho C-S, et al. Nomenclature for factors of the SLA class-I system, 2004. Tissue Antigens. 2005;65:136–49.

23. Martens GW, Lunney JK, Baker JE, Smith DM. Rapid assignment of swine leukocyte antigen haplotypes in pedigreed herds using a polymerase chain reaction-based assay. Immunogenetics. 2003;55:395–401.

24. Ho C-S, Lunney JK, Franzo-Romain MH, Martens GW, Lee Y-J, Lee J-H, et al. Molecular characterization of swine leucocyte antigen class I genes in outbred pig populations. Anim. Genet. 2009;40:468–78.

25. Ho C-S, Lunney JK, Lee J-H, Franzo-Romain MH, Martens GW, Rowland RRR, et al. Molecular characterization of swine leucocyte antigen class II genes in outbred pig populations. Anim. Genet. 2010;41:428–32.

26. Li Pira G, Ivaldi F, Moretti P, Manca F. High throughput T epitope mapping and vaccine development. J. Biomed. Biotechnol. 2010;2010:325720.

27. Ober BT, Summerfield A, Mattlinger C, Wiesmüller KH, Jung G, Pfaff E, et al. Vaccine-induced, pseudorabies virus-specific, extrathymic CD4+CD8+ memory T-helper cells in swine. J. Virol. 1998;72:4866–73.

28. Blanco E, McCullough K, Summerfield A, Fiorini J, Andreu D, Chiva C, et al. Interspecies major histocompatibility complex-restricted Th cell epitope on foot-and-mouth disease virus capsid protein VP4. J. Virol. 2000;74:4902–7.

29. Blanco E, Garcia-Briones M, Sanz-Parra A, Gomes P, De Oliveira E, Valero ML, et al. Identification of T-cell epitopes in nonstructural proteins of foot-and-mouth disease virus. J. Virol. 2001;75:3164–74.

30. García-Briones MM, Blanco E, Chiva C, Andreu D, Ley V, Sobrino F. Immunogenicity and T cell recognition in swine of foot-and-mouth disease virus polymerase 3D. Virology. 2004;322:264–75.

31. Stevenson LS, Gilpin DF, Douglas A, McNeilly F, McNair I, Adair BM, et al. T lymphocyte epitope mapping of porcine circovirus type 2. Viral Immunol. 2007;20:389–98.

32. Armengol E, Wiesmüller K-H, Wienhold D, Büttner M, Pfaff E, Jung G, et al. Identification of T-cell epitopes in the structural and non-structural proteins of classical swine fever virus. J. Gen. Virol. 2002;83:551–60.

33. Gerner W, Denyer MS, Takamatsu H-H, Wileman TE, Wiesmüller K-H, Pfaff E, et al. Identification of novel foot-and-mouth disease virus specific T-cell epitopes in c/c and d/d haplotype miniature swine. Virus Res. 2006;121:223–8.

34. Vashisht K, Goldberg TL, Husmann RJ, Schnitzlein W, Zuckermann FA. Identification of immunodominant T-cell epitopes present in glycoprotein 5 of the North American genotype of porcine reproductive and respiratory syndrome virus. Vaccine. 2008;26:4747–53.

35. Wang Y-X, Zhou Y-J, Li G-X, Zhang S-R, Jiang Y-F, Xu A-T, et al. Identification of immunodominant T-cell epitopes in membrane protein of highly pathogenic porcine reproductive and respiratory syndrome virus. Virus Res. 2011;158:108–15.

36. Parida R, Choi I-S, Peterson DA, Pattnaik AK, Laegreid W, Zuckermann FA, et al. Location of T-cell epitopes in nonstructural proteins 9 and 10 of type-II porcine reproductive and respiratory syndrome virus. Virus Res. 2012;169:13–21.

37. Im E-J, Hong JP, Roshorm Y, Bridgeman A, Létourneau S, Liljeström P, et al. Protective efficacy of serially up-ranked subdominant CD8+ T cell epitopes against virus challenges. PLoS Pathog. 2011;7:e1002041.

38. Saade F, Gorski SA, Petrovsky N. Pushing the frontiers of T-cell vaccines: accurate measurement of human T-cell responses. Expert Rev. Vaccines. 2012;11:1459–70.

39. Pedersen LE, Harndahl M, Rasmussen M, Lamberth K, Golde WT, Lund O, et al. Porcine major histocompatibility complex (MHC) class I molecules and analysis of their peptide-binding specificities. Immunogenetics. 2011;63:821–34.

40. Pedersen LE, Jungersen G, Buus S, Golde WT. Analysis of Swine Leukocyte Antigen peptide binding profiles and the identification of T cell epitopes by tetramer staining [Ph.D. thesis]. Technical University of Denmark; 2012.

41. Pedersen LE, Harndahl M, Nielsen M, Patch JR, Jungersen G, Buus S, et al. Identification of peptides from foot-and-mouth disease virus structural proteins bound by class I swine leukocyte antigen (SLA) alleles, SLA-1*0401 and SLA-2*0401. Anim. Genet. 2013;44:251–8.

42. Pedersen LE, Rasmussen M, Harndahl M, Nielsen M, Buus S, Jungersen G. A combined prediction strategy increases identification of peptides bound with high affinity and stability to porcine MHC class I molecules SLA-1*04:01, SLA-2*04:01, and SLA-3*04:01. Immunogenetics. 2016;68:157–65.

43. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The immune epitope database (IEDB) 3.0. Nucleic Acids Res. 2015;43:D405-412.

44. Zhang L, Udaka K, Mamitsuka H, Zhu S. Toward more accurate pan-specific MHC-peptide binding prediction: a review of current methods and tools. Brief. Bioinform. 2012;13:350–64.

45. Madden DR. The three-dimensional structure of peptide-MHC complexes. Annu. Rev. Immunol. 1995;13:587–622.

46. Sinigaglia F, Hammer J. Defining rules for the peptide-MHC class II interaction. Curr. Opin. Immunol. 1994;6:52–6.

47. Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U, et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat. Biotechnol. 1999;17:555–61.

48. Zhang L, Chen Y, Wong H-S, Zhou S, Mamitsuka H, Zhu S. TEPITOPEpan: Extending TEPITOPE for Peptide Binding Prediction Covering over 700 HLA-DR Molecules. PLoS ONE. 2012;7:e30483.

49. Zhang H, Lund O, Nielsen M. The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. Bioinformatics. 2009;25:1293–9.

50. De Groot AS, Jesdale BM, Szu E, Schafer JR, Chicz RM, Deocampo G. An interactive Web site providing major histocompatibility ligand predictions: application to HIV research. AIDS Res. Hum. Retroviruses. 1997;13:529–31.

51. Gregory SH, Mott S, Phung J, Lee J, Moise L, McMurry JA, et al. Epitope-based vaccination against pneumonic tularemia. Vaccine. 2009;27:5299–306.

52. Moise L, Buller RM, Schriewer J, Lee J, Frey SE, Weiner DB, et al. VennVax, a DNA-prime, peptide-boost multi-T-cell epitope poxvirus vaccine, induces protective immunity against vaccinia infection by T cell response alone. Vaccine. 2011;29:501–11.

53. Moise L, Tassone R, Latimer H, Terry F, Levitz L, Haran JP, et al. Immunization with cross-conserved H1N1 influenza CD4+ T-cell epitopes lowers viral burden in HLA DR3 transgenic mice. Hum. Vaccines Immunother. 2013;9:2060–8.

54. Zhao L, Zhang M, Cong H. Advances in the study of HLA-restricted epitope vaccines. Hum. Vaccines Immunother. 2013;9:2566–77.

55. Xiao Y-F, Jie M-M, Li B-S, Hu C-J, Xie R, Tang B, et al. Peptide-Based Treatment: A Promising Cancer Therapy. J. Immunol. Res. 2015;2015:761820.

56. Moss SF, Moise L, Lee DS, Kim W, Zhang S, Lee J, et al. HelicoVax: epitope-based therapeutic Helicobacter pylori vaccination in a mouse model. Vaccine. 2011;29:2085–91.

57. Mallone R, Nepom GT. MHC Class II tetramers and the pursuit of antigen-specific T cells: define, deviate, delete. Clin. Immunol. Orlando Fla. 2004;110:232–42.

58. Moise L, Song C, Martin WD, Tassone R, De Groot AS, Scott DW. Effect of HLA DR epitope de-immunization of Factor VIII in vitro and in vivo. Clin. Immunol. 2012;142:320–31.

59. Zhao H, Verma D, Li W, Choi Y, Ndong C, Fiering SN, et al. Depletion of T cell epitopes in lysostaphin mitigates anti-drug antibody response and enhances antibacterial efficacy in vivo. Chem. Biol. 2015;22:629–39.

60. Otten HG, Calis JJA, Keşmir C, van Zuilen AD, Spierings E. Predicted indirectly recognizable HLA epitopes presented by HLA-DR correlate with the de novo development of donor-specific HLA IgG antibodies after kidney transplantation. Hum. Immunol. 2013;74:290–6.

61. Prickett SR, Rolland JM, O'Hehir RE. Immunoregulatory T cell epitope peptides: the new frontier in allergy therapy. Clin. Exp. Allergy. 2015;45:1015–26.

62. Inaba H, De Groot LJ, Akamizu T. Thyrotropin Receptor Epitope and Human Leukocyte Antigen in Graves' Disease. Front. Endocrinol. 2016;7.

63. Nandy A, Basak SC. A Brief Review of Computer-Assisted Approaches to Rational Design of Peptide Vaccines. Int. J. Mol. Sci. 2016;17.

64. Zhang Q, Wang P, Kim Y, Haste-Andersen P, Beaver J, Bourne PE, et al. Immune epitope database analysis resource (IEDB-AR). Nucleic Acids Res. 2008;36:W513-518.

65. Nielsen M, Andreatta M. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. Genome Med. 2016;8:33.

66. Donovan TS. The role of influenza on growing pig performance. Allen Leman Swine Conf. Saint Paul, MN; 2005. p. 97–8.

67. Wideman G. Impact of Influenza A in Pork Production. Proc. Lond. Swine Conf. London, Ontario; 2013. p. 191–3.

68. Lorusso A, Vincent AL, Gramer ME, Lager KM, Ciacci-Zanella JR. Contemporary epidemiology of North American lineage triple reassortant influenza A viruses in pigs. Curr. Top. Microbiol. Immunol. 2013;370:113–32.

69. Vincent AL, Lager KM, Anderson TK. A brief introduction to influenza A virus in swine. Methods Mol. Biol. Clifton NJ. 2014;1161:243–58.

70. Anderson TK, Campbell BA, Nelson MI, Lewis NS, Janas-Martindale A, Killian ML, et al. Characterization of co-circulating swine influenza A viruses in North America and the identification of a novel H1 genetic clade with antigenic significance. Virus Res. 2015;201:24–31.

71. Ma W, Kahn RE, Richt JA. The pig as a mixing vessel for influenza viruses: Human and veterinary implications. J. Mol. Genet. Med. Int. J. Biomed. Res. 2008;3:158–66.

72. Kreijtz JHCM, Fouchier RAM, Rimmelzwaan GF. Immune responses to influenza virus infection. Virus Res. 2011;162:19–30.

73. Brandenburg B, Koudstaal W, Goudsmit J, Klaren V, Tang C, Bujny MV, et al. Mechanisms of hemagglutinin targeted influenza virus neutralization. PloS One. 2013;8:e80034.

74. Lee LY-H, Ha DLA, Simmons C, de Jong MD, Chau NVV, Schumacher R, et al. Memory T cells established by seasonal human influenza A infection cross-react with avian influenza A (H5N1) in healthy individuals. J. Clin. Invest. 2008;118:3478–90.

75. Price GE, Soboleski MR, Lo C-Y, Misplon JA, Pappas C, Houser KV, et al. Vaccination focusing immunity on conserved antigens protects mice and ferrets against virulent H1N1 and H5N1 influenza A viruses. Vaccine. 2009;27:6512–21.

76. Babon JAB, Cruz J, Orphin L, Pazoles P, Co MDT, Ennis FA, et al. Genome-wide screening of human T-cell epitopes in influenza A virus reveals a broad spectrum of CD4(+) T-cell responses to internal proteins, hemagglutinins, and neuraminidases. Hum. Immunol. 2009;70:711–21.

77. La Gruta NL, Turner SJ. T cell mediated immunity to influenza: mechanisms of viral control. Trends Immunol. 2014;35:396–402.

78. Altenburg AF, Rimmelzwaan GF, de Vries RD. Virus-specific T cells as correlate of (cross-)protective immunity against influenza. Vaccine. 2015;33:500–6.

79. Lin YL, Askonas BA. Biological properties of an influenza A virus-specific killer T cell clone. Inhibition of virus replication in vivo and induction of delayed-type hypersensitivity reactions. J. Exp. Med. 1981;154:225–34.

80. McMichael AJ, Gotch FM, Noble GR, Beare PA. Cytotoxic T-cell immunity to influenza. N. Engl. J. Med. 1983;309:13–7.

81. Khatri M, Dwivedi V, Krakowka S, Manickam C, Ali A, Wang L, et al. Swine Influenza H1N1 Virus Induces Acute Inflammatory Immune Responses in Pig Lungs: a Potential Animal Model for Human H1N1 Influenza Virus. J. Virol. 2010;84:11210–8.

82. Kitikoon P, Nilubol D, Erickson BJ, Janke BH, Hoover TC, Sornsen SA, et al. The immune response and maternal antibody interference to a heterologous H1N1 swine influenza virus infection following vaccination. Vet. Immunol. Immunopathol. 2006;112:117–28.

83. Loving CL, Vincent AL, Pena L, Perez DR. Heightened adaptive immune responses following vaccination with a temperature-sensitive, live-attenuated influenza virus compared to adjuvanted, whole-inactivated virus in pigs. Vaccine. 2012;30:5830–8.

84. Platt R, Vincent AL, Gauger PC, Loving CL, Zanella EL, Lager KM, et al. Comparison of humoral and cellular immune responses to inactivated swine influenza virus vaccine in weaned pigs. Vet. Immunol. Immunopathol. 2011;142:252–7.

85. Sandbulte MR, Platt R, Roth JA, Henningson JN, Gibson KA, Rajão DS, et al. Divergent immune responses and disease outcomes in piglets immunized with inactivated and attenuated H3N2 swine influenza vaccines in the presence of maternally-derived antibodies. Virology. 2014;464–465:45–54.

86. Sandbulte MR, Spickler AR, Zaabel PK, Roth JA. Optimal Use of Vaccines for Control of Influenza A Virus in Swine. Vaccines. 2015;3:22–73.

87. Vincent AL, Ciacci-Zanella JR, Lorusso A, Gauger PC, Zanella EL, Kehrli Jr. ME, et al. Efficacy of inactivated swine influenza virus vaccines against the 2009 A/H1N1 influenza virus in pigs. Vaccine. 2010;28:2782–7.

88. Gramer M, Rossow K. Epidemiology of swine influenza and implications of reassortment. Allen Leman Swine Conf. 2004. p. 69–73.

89. Rapp-Gabrielson VJ, Lenz M, Hildebrand T, Taylor L, Kuhn M. Evaluation of cross-protection of FluSure XP® against a heterologous gamma cluster H1N1 swine influenza virus challenge. Allen Leman Swine Conf. 2011. p. 267.

90. Rapp-Gabrielson VJ, Lenz M, Kuhn M, Taylor L, Culhane M, Kesl L, et al. Cross-protection of FluSure® XP in pigs challenged with a Gamma cluster H1N1/pH1N1 reassortant swine influenza virus. Allen Leman Swine Conf. 2013. p. 219.

91. Detmer SE, Gramer MR, King VL, Mathur S, Rapp-Gabrielson VJ. In vivo evaluation of vaccine efficacy against challenge with a contemporary field isolate from the α cluster of H1N1 swine influenza virus. Can. J. Vet. Res. Rev. Can. Rech. Vét. 2013;77:24–32.

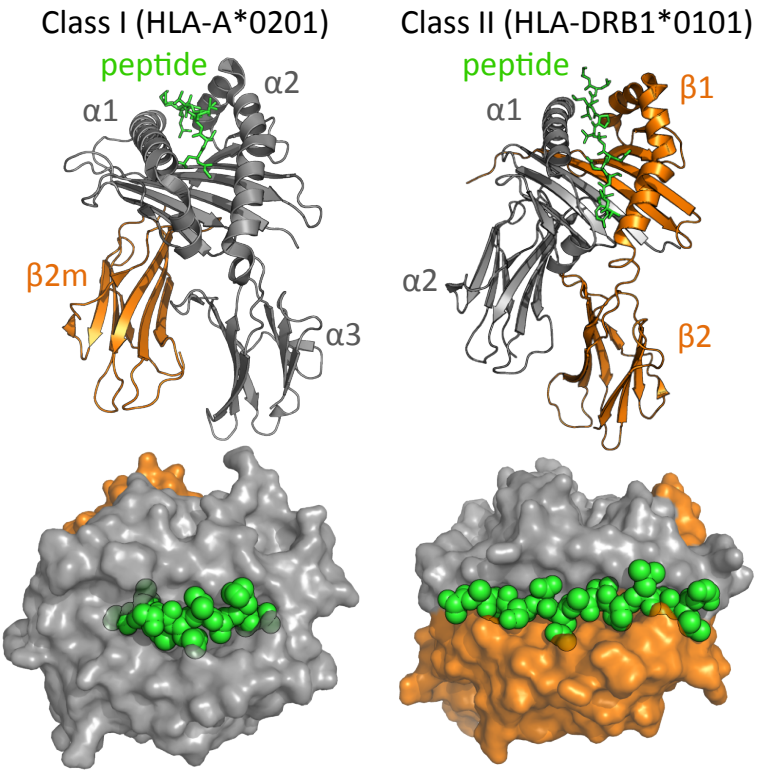


Fig. 1. Comparison of MHC class I and class II crystal structures. (Top; side view) Ribbon diagrams of class I (HLA-A*0201; PDB id: 3MRE) and class II (HLA-DRB1*0201; PDB id: 1T5W) MHC proteins. Grey regions are α-domains, orange regions β-domains. The transmembrane and short cytoplasmic domains, which are not visualized in the crystal structures, extend towards the bottom of the figure. Peptides (green) bind above the large β-sheet and between the α-helical regions. The domain organization of class I (α1, α2, α3, β2m) and class II (α1, α2, β1, β2) is different; however, their three dimensional structures are very similar. (Bottom; top view) Surface representations of MHC molecules are shown and peptides are represented as spheres. Small conformational alterations and amino acid substitutions in the peptide-binding site account for the difference in peptide length preference between class I (8–10mers) and class II (>12mers). The class I peptide has its ends buried in the binding site, and arches away from the HLA protein in the center. The class II peptide is straight, and extends out of the open ends of the groove.

CHAPTER 2

Manuscript published in BMC Bioinformatics, September 2015

Development and validation of an epitope prediction tool for swine (PigMatrix) based on the pocket profile method

Andres H. Gutiérrez[1], William D. Martin[2], Chris Bailey-Kellogg[3], Frances Terry[2], Leonard Moise[1,2], Anne S. De Groot[1,2]*

[1]Institute for Immunology and Informatics, Department of Cell and Molecular Biology, University of Rhode Island, Providence, RI, USA

[2]EpiVax, Inc., Providence, RI, USA

[3]Department of Computer Science, Dartmouth College, Hanover, NH, USA

*Corresponding Author: Anne S. De Groot, M.D.
Institute for Immunology and Informatics
Department of Cell and Molecular Biology
University of Rhode Island
Providence, RI, 02903, USA
E-mail: dr.annie.degroot@gmail.com

Abstract

Background: T cell epitope prediction tools and associated vaccine design algorithms have accelerated the development of vaccines for humans. Predictive tools for swine and other food animals are not as well developed, primarily because the data required to develop the tools are lacking. Here, we overcome a lack of T cell epitope data to construct swine epitope predictors by systematically leveraging available human information. Applying the "pocket profile method", we use sequence and structural similarities in the binding pockets of human and swine major histocompatibility complex proteins to infer Swine Leukocyte Antigen (SLA) peptide binding preferences.

Methods: We developed epitope-prediction matrices (PigMatrices), for three SLA class I alleles (SLA-1*0401, 2*0401 and 3*0401) and one class II allele (SLA-DRB1*0201), based on the binding preferences of the best-matched Human Leukocyte Antigen (HLA) pocket for each SLA pocket. The contact residues involved in the binding pockets were defined for class I based on crystal structures of either SLA (SLA-specific contacts, Ssc) or HLA supertype alleles (HLA contacts, Hc); for class II, only Hc was possible. Different substitution matrices were evaluated (PAM and BLOSUM) for scoring pocket similarity and identifying the best human match. The accuracy of the PigMatrices was compared to available online swine epitope prediction tools such as PickPocket and NetMHCpan.

Results: PigMatrices that used Ssc to define the pocket sequences and PAM30 to score pocket similarity demonstrated the best predictive performance and were able to accurately separate binders from random peptides. For SLA-1*0401 and 2*0401, PigMatrix achieved area under the receiver operating characteristic curves (AUC) of 0.78 and 0.73, respectively, which were equivalent or better than PickPocket (0.76 and 0.54) and NetMHCpan version 2.4 (0.41 and 0.51) and version 2.8 (0.72 and 0.71). In addition, we developed the first predictive SLA class II matrix, obtaining an AUC of 0.73 for existing SLA-DRB1*0201 epitopes. Notably, PigMatrix achieved this level of predictive power without training on SLA binding data.

Conclusion: Overall, the pocket profile method combined with binding preferences from HLA binding data shows significant promise for developing T cell epitope prediction tools for pigs. When combined with existing vaccine design algorithms, PigMatrix will be useful for developing genome-derived vaccines for a range of pig pathogens for which no effective vaccines currently exist (e.g. porcine

reproductive and respiratory syndrome, influenza and porcine epidemic diarrhea).

Keywords: PigMatrix, EpiMatrix, Computational vaccinology, Epitope prediction, HLA, SLA, MHC, class I, class II, Porcine, PRRSV, Influenza, Genome-derived vaccine, T cell epitope

Background

The interaction of Major Histocompatibility Complex (MHC) proteins with peptides derived from protein antigens plays a key role in the adaptive immune response mediated by T cells. The MHC:peptide complex presented on the surface of a cell is recognized by the T cell receptor (TCR), which activates the T cell and drives the immune response. There are two classes of MHC molecules: MHC class I presents peptides of intracellular origin to $CD8^+$ T cells (cytotoxic T cells, or CTL) and MHC class II presents peptides of extracellular origin to $CD4^+$ T cells (T-helper cells, or Th). Both classes of molecules have similar tertiary structure. Class I molecules have a transmembrane (α) chain noncovalently associated with β_2-microglobulin where the α_1 and α_2 domains form the peptide-binding groove; class II molecules have two transmembrane chains (α and β) where the α_1 and β_1 domains form the peptide-binding groove. The MHC class I binding groove is closed, which restricts the length of bound peptides to 8-10 residues; the MHC class II binding groove on the other hand, is open, and peptides can extend beyond the ends of the groove, allowing binding of longer and more flexible peptides of variable lengths (typically 13-25 amino acids) [1].

The tertiary structure of MHC molecules is relatively conserved, even across species. For example, crystallographic studies have shown similarity between Human Leukocyte Antigen (HLA; human MHC) and Swine Leukocyte Antigen (SLA; swine MHC) molecules [2]. The SLA-1*0401 class I allele has been crystallized in complex with peptides derived from 2009-pandemic H1N1 (pH1N1) swine-origin influenza A virus and Ebola virus. A structural comparison revealed that the SLA class I molecule, SLA-1*0401 contains six pockets in its binding groove, similar to HLA class I molecules. The root-mean-squared deviation (RMSD) for all of the Cα atoms in SLA-1*0401 and HLA-A*1101, which has the highest identity with SLA-1*0401 (78%), was <0.7Å indicating a similar arrangement of their backbones. Furthermore, three out of 23 influenza SLA-1*0401 binders were identical to previously defined peptides presented by HLA-A*0101 [2]. For SLA class II, no crystal structures are available, but amino acid SLA-DR sequences are highly similar to their human counterparts. For example, the amino acid sequences of SLA-DRB1*0201 and HLA-DRB1*0101 are 79% identical.

Due to the importance of peptide binding to MHC molecules in the immune response, human T cell epitope prediction tools have been developed based on a range of approaches and are widely used

in vaccine development and experimental immunology [3]. The availability of a large and expanding database of validated MHC ligands has contributed to the development of more accurate algorithms. Epitope predictions using these tools reduce the time and effort required to identify T cell epitopes [4]. The number of epitope prediction tools is more limited for pigs due to the paucity of experimental data available. To overcome the lack of quantitative measurements of MHC interaction for a large number of HLA alleles, 'pan-specific' methods have been implemented for prediction of T cell epitopes that bind to MHC for which experimental data are limited or not available. Pan-specific methods use experimental binding data and amino acid sequences of multiple MHC alleles to infer binding preferences to uncharacterized MHC molecules. These methods have been used for development of prediction tools for MHC class I [5–7] and II alleles [8–11], but only NetMHCpan has been used for prediction of SLA class I-restricted peptides [12–16]. This method is based on artificial neural networks (ANN) trained using as input a pseudo-sequence composed of the polymorphic residues in the binding groove of a given MHC, a peptide sequence and the experimental affinity data. To our knowledge, there are no in silico tools that are available for SLA class II.

Sturniolo et al. first described a method for using existing data to develop new epitope predictors, the pocket profile method, in 1999 [17]. It has been used to develop pan-specific methods for predicting binding of peptides to HLA class I and II alleles [9, 18]. The approach depends on the identification of certain polymorphic regions within HLA molecules that are known to be the areas of contact between peptides and the binding groove of HLA [19–21]. Contact residues from the HLA molecule that bind the R group (side chain) of a specific amino acid within a linear peptide can be considered to form a pocket for that R group. Thus, each 'pocket' can also be described in terms of its amino acid binding preferences ('pocket profile'). The pocket profiles are nearly independent of the remaining binding groove. So, this method assumes that two MHC alleles with identical pocket residues will have the same pocket profile. Therefore, given sufficient information about the contact residues of the set of pockets in the binding groove of an MHC and experimentally determined pocket profiles, it is possible to compose predictive matrices in silico. The method was originally applied to develop TEPITOPE, an algorithm for prediction of peptide ligands to 51 HLA class II alleles with known pocket residues [17] and then extended to any HLA-DR molecules with similar pockets (TEPITOPEpan) [9]. A

similar method has also been used in the PickPocket algorithm for MHC class I prediction [18]. Whereas TEPITOPEpan uses pocket profiles from TEPITOPE, PickPocket generates binding preferences using position-specific scoring matrices (PSSMs) from binding data directly. Although no publications exist using these algorithms for SLA binding predictions, SLA alleles are available for use in the PickPocket server (www.cbs.dtu.dk/services/PickPocket/).

EpiMatrix is a matrix-based algorithm that uses the pocket profile method to predict potential HLA class I and II T cell epitopes. The first version of this algorithm was developed in 1996, and newer versions have been extensively validated in vitro in HLA binding assay and human T cell assays and in animal studies using HLA transgenic mouse models [22–26]. For common class II alleles, EpiMatrix appears to predict more accurately than many available epitope-mapping algorithms [27]. Comparative performance for EpiMatrix class I predictions has not been published; however, the tools have been successfully applied to identify class I-restricted T cell epitopes in human pathogens [28–30]. The pocket profile method was used to develop a matrix for a bovine MHC class I allele [31] and in the early 2000s, this method was also used to derive SLA class II prediction matrices from EpiMatrix, but this work was not published.

This paper describes the development and retrospective validation of predictive matrices to map T cell epitopes for SLA class I (SLA-1*0401, 2*0401, 3*0401) alleles and a class II (SLA-DRB1*0201) allele. “PigMatrix” matrices are built based on the pocket profile method using EpiMatrix pocket profiles for HLA epitope prediction. While these alleles represent a small subset of commonly expressed alleles in pigs [32–36], they were selected for their available quantitative peptide data [13–16, 37–39]. As before, we assumed that predictive matrices developed for HLA alleles should function as reasonable proxies for the prediction of ligands to SLA molecules with similar pocket profiles. Thus, we developed ‘composite matrices’ by selecting the most similar HLA pocket (best human match) for each SLA pocket, and built matrices composed of the corresponding HLA binding preferences (Fig. 1).

Two methods were used to define the pocket contact residues considering different scenarios of availability of SLA crystal structures. In the first scenario, SLA crystal structures were available, so pockets were defined from these structures. In the second, no SLA crystal structures were available; therefore, contact residues were selected based on crystal structures of HLA. We also tested different

substitution matrices (PAM and BLOSUM) to score pocket similarity to define the best human match. PigMatrix was benchmarked against existing SLA prediction tools for class I alleles. Benchmarking against other SLA class II predictors was not possible as no other prediction algorithms are available. The results demonstrate the potential of this approach to develop matrices to make accurate predictions for both SLA class I and II alleles for which experimental binding data are limited or even non-existent.

Methods

Datasets

Unique 9-mer peptides with reported binding measurement to a specific SLA allele were compiled from the literature into two datasets: one comprising binders and the other, non-binders, for each of three class I (SLA-1*0401, 2*0401, 3*0401) alleles [2, 13–16]. The SLA-1*0401 dataset included 133 binders and 46 non-binders; 2*0401 included 24 binders and 46 non-binders; and 3*0401, 27 binders and 46 non-binders. Twenty-five (14%) of the SLA-1*0401 peptides were reported by Zhang et al. [2]; the remaining peptides for 1*0401, 2*0401 and 3*0401 were published by Pedersen et al. in different publications [13–16]. For class II, a dataset was created with peptides specific to SLA-DRB1*0201 from the literature [37–39]. This dataset has 33 binders and 171 non-binders. Peptides with contradictory (both positive and negative) results were discarded (Supplemental Table 1). Additionally, we generated a set of 100,000 unique 9-mer peptides from random sequence proteins with the average amino acid frequencies of the proteins in the Swiss-Prot database for use as a control data set, as previously described [31]. The random proteins were generated using the RandSeq tool from ExPASy [40].

MHC sequences

Complete amino acid sequences from SLA proteins, along with HLA class I (HLA-A*0101, A*0201, A*0301, A*1101, A*2402, A*6801, B*0702, B*0801, B*2705, B*3501, B*4403, B*5101) and class II alleles (HLA-DRB1*0101, 0301, 0401, 0701, 0801, 1101, 1301, 1501), were obtained from the IPD-MHC Database (www.ebi.ac.uk/ipd/). It is important to clarify that the HLA alleles for which binding preferences are available in EpiMatrix are families of alleles that share pocket preferences, rather than individual alleles. The alleles represent 12 class I supertypes [41] and eight class II supertypes [42].

Binding pocket residues

Six pockets (A-F) and five pockets (A-E) were considered for class I and II, respectively. Pockets for peptide positions 4, 5 and 8 for class I and 2, 3, 5, and 8 for class II were not considered due to their minimal effect on binding [17, 19, 43]. For each pocket, contact residues (pocket sequences) were defined as either (1) SLA-specific contacts (Ssc) derived from SLA crystal structures or (2) HLA-based contacts (Hc) derived from HLA crystal structures (Fig. 2A). The Ssc approach was applied only to SLA class I alleles using crystallographic data available for SLA-1*0401 (PDB:3QQ3 and 3QQ4) [2]. For Hc, representative crystal structures from HLA class I and II supertype alleles [41, 42] with bound 9-mer (for class I) or longer peptides (for class II) and the highest resolution were analyzed to define the contact residues (Supplemental Table 2). Four class II supertype alleles (HLA-DRB1*0101, 0301, 0401, and 1501) had crystal structures available.

We considered binding pocket contact residues to be amino acids with atoms within 5.0Å of those in the bound peptide. Residues were selected using PyMOL (Schrodinger, LLC). Only amino acids with the side chain oriented towards the peptide were included. Thus, each pocket included the union of contact residues in all the MHC crystal structures of a given class. For class II alleles, since the alpha subunit of HLA-DR (HLA-DRA) is practically invariable [44], only residues in the beta subunit were included. The amino acids in the positions defined as contact residues according to Ssc and Hc were extracted from HLA sequences and compiled into a pocket library (Fig. 2B), where each pocket is a non-contiguous sequence of residues ordered by their positions.

Composite matrix construction

Each SLA protein sequence was aligned to a reference HLA sequence (HLA-A*0101 for class I and HLA-DRB1*0101 for class II) to extract its contact residues (Fig. 2C) based on Ssc and Hc approaches. SLA pockets were compared to the HLA pocket library to identify the best human match for each SLA pocket. SLA-HLA pocket similarity was determined using PAM and BLOSUM substitution matrices for closely (PAM30 and BLOSUM90) and distantly related (PAM120 and BLOSUM62) protein sequences [45, 46]. For a pocket comparison between SLA sequence x and corresponding HLA sequence y, both of length N, the similarity score was calculated as the sum of the similarity scores of each amino acid i using a specific substitution matrix M. The score was then divided by the similarity

score of the SLA pocket compared to itself.

$$\mathrm{sim}_{\mathrm{M}}(x, y) = \frac{\sum_{i=1}^{N} \mathrm{M}[x_i, y_i]}{\sum_{i=1}^{N} \mathrm{M}[x_i, x_i]}$$

The HLA pocket with the highest similarity score was considered to be the best human match (Fig. 2D). The pocket profiles of the best human matches for each pocket were then combined to form composite matrices (Fig. 2E).

Matrix validation and performance evaluation

Composite matrices were used to score a set of random 9-mer peptides. The raw binding score $\mathrm{bind}_{\mathrm{raw}}(\mathrm{p})$ for each peptide p was calculated as the sum, over a set i of relevant peptide positions, of the coefficient $K[i,p_i]$ of the amino acid p_i at position i in p. Positions 1, 2, 3, 6, 7, and 9 were used for class I and 1, 4, 6, 7, and 9 for class II, as those positions most interact with each SLA pocket.

$$\mathrm{bind}_{\mathrm{raw}}(\mathrm{p}) = \sum_{i \in I} K[i, p_i]$$

The average μ and standard deviation σ of the scores were used to normalize scores into a Z-score scale (binding likelihood score).

$$Z = \frac{\mathrm{bind}_{\mathrm{raw}}(\mathrm{p}) - \mu}{\sigma}$$

Next, the ability of the composite matrices to separate binders from non-binders and binders from a set of random peptides was evaluated by comparing the mean of the Z-scores of the datasets [28]. Differences were evaluated for significance by Wilcoxon-Mann–Whitney test. For class II binders longer than 10 amino acids, 9-mers overlapping by eight amino acids were scored because in general, the lengths of MHC binding cores are 9 amino acids [47]. The 9-mer frame with the highest Z-score was selected to be the most likely MHC binder and its score was used for calculation of the mean Z-score of binders and non-binders. In addition, for each allele, the HLA matrix with the lowest overall pocket identity with the SLA was used to score both set of peptides, binders and random peptides, as a negative control matrix.

Peptides in the top 5% of the normal curve, where the Z-score is greater than or equal to 1.64, were considered to be potential binders. This threshold has been shown to identify peptides that are highly likely to bind HLA molecules [22]. So as to evaluate the predictive performance of the matrices,

we calculated the area under the receiver operating characteristic (ROC) curve (AUC) using the *sensitivity* and *1 – specificity* values for the same dataset of binders and non-binders.

Finally, PigMatrix SLA class I predictions were compared to those of PickPocket 1.1 and NetMHCpan 2.4 and 2.8. A threshold of 500 nM in binding affinity (or 0.426 prediction score based on $1 - log_{50K}(affinity)$), was set to classify binders and non-binders as previously described [6, 18].

Results

Pocket residues

The contact residues that form the binding pockets in SLA were defined from (1) SLA-1*0401 crystal structures (SLA-specific contacts, Ssc), and (2) HLA class I and II crystal structures (HLA contacts, Hc). Fig. 3 shows contact residue similarities and differences for SLA-1*0401 in determinations using Hc and Ssc. Thirty-nine positions were identified with Hc, of which 34 were in common with Ssc (shown in light blue in Fig. 3) and five were unique to Hc (in orange); there were no positions unique to Ssc. Several amino acids were involved in more than one pocket; however, this was more frequent for Hc than Scc due to the nature of the approach; only 23 of the 34 common positions belonged to exactly the same pockets by both definitions (positions shown in bold and underlined in Fig. 3). Hc included for each pocket, the union of amino acids over all the HLA crystal structures analyzed. Based on Hc, positions 97, 99 and 114 were part of four pockets; these residues are located in the central part of the MHC binding groove and depending on the characteristics of their R chain and the bound peptide, they can interact with more than one residue of the ligand. We also observed that the total number of contact residues per pocket was lower in Ssc. The main differences in SLA class I were observed for pockets C and D where SLA structures had fewer contact residues involved in the binding. Pocket F, on the other hand, was identical for both.

For class II, only Hc was applied because no SLA-DR crystal structures were available. Twenty-two positions were considered in contact with ligands; four positions were common to three pockets, four to two pockets and 14 were involved in only one pocket. Five positions were included for pocket A, seven in B, seven in C, eight in D, and seven in F (Supplemental Table 3). In pockets B, C, D, and E, we identified in total eight allele-specific pocket residues; Y30 in pocket B for HLA-

DRB1*1501, D28 and R74 in pocket C for DRB1*0301, Q70 and R74 for DRB1*0301 and R13 for DRB1*1501 in pocket D, and V38 for DRB1*0301 and Y37 for DRB1*0401 in pocket E.

Matrix construction, validation and evaluation

We built composite matrices for each SLA allele and evaluated whether they were able to distinguish SLA allele-specific binders from non-binders and random peptides (Fig. 4). In some cases, the best human matches were the same regardless of the approach used to define pocket residues and the pocket similarity scoring method applied; therefore, prediction results were identical (e.g. Fig. 4 top, SLA-2*0401). For SLA-1*0401 and 2*0401, two and four scoring methods respectively, generated Ssc matrices capable of separating binders from non-binders (highlighted in gray in Fig. 4 top). For these matrices, mean Z-scores of binders were above the threshold to be considered a potential binder (1.64) and non-binder Z-scores were below. Furthermore, the difference between the sets of peptides was statistically significant ($p<0.001$) using a Wilcoxon-Mann-Whitney test. None of the class I Hc matrices was able to distinguish with statistical significance binders from non-binders. Likewise, for all SLA-3*0401 matrices, mean Z-scores of the binders were either not above the 1.64 threshold or the non-binders had higher mean Z-scores. For class II allele SLA-DRB1*0201, binders scored using Hc matrices were above the threshold and were statistically distinct from mean Z-scores of the non-binders ($p<0.01$). Negative control matrices for all SLA alleles (using HLA alleles with the lowest overall pocket identity) did not separate binders from random peptides (means range from −0.57 to 1.13), showing that the selection of the best human match based on similarity is critical. In sum, these results show that some composite class I Ssc matrices and class II Hc were able to separate binders from random peptides and non-binders.

To evaluate the predictive performance of the matrices, we built ROC curves and then calculated the AUC. For the AUC, a value of 0.5 corresponds to a random prediction and a value of 1 to a perfect prediction. Fig. 5 shows a comparison of the AUCs of Ssc and Hc matrices, PickPocket and NetMHCpan. For class I, matrices based on Ssc had higher AUCs than Hc-based matrices. Class I and II matrices constructed using PAM30 to score pocket similarity had higher AUCs compared to matrices constructed using PAM120, BLOSUM62 and PAM120, with one exception (Hc SLA-1*0401 built using BLOSUM62). Compared to PickPocket and NetMHCpan 2.4 and 2.8, PigMatrix's AUC was

equivalent or better for SLA-1*0401 and 2*0401; however, due to the nature of the tests, we could not assess statistical significance. It is worth noting that, in contrast with NetMHCpan 2.8, SLA peptide data did not contribute to training PigMatrix.

The SLA-3*0401 PAM30-Ssc matrix had the lowest AUC (0.59). This was not unexpected as matrices for this allele were unable to separate binders from non-binders as described above (Fig. 4). For these reasons, we examined the binding preferences of the best human matches for the PAM30-Ssc matrix and compared them to the amino acid frequencies in the sets of binders and non-binders. The most evident differences were observed in pocket B. The best human match for pocket B was HLA-A*0301 (sim_{PAM30} 0.49). Of all binders reported for SLA-3*0401, the most common residue in position two (Pocket B) was arginine, found in 37% of binders, followed by alanine, found in 19% of binders. These frequencies were more similar to the binding preferences in HLA-B*2705 (sim_{PAM30} 0.17) pocket profile, in contrast to A*0301, which had negative coefficients for arginine and alanine. Based on this observation, we modified pocket B of PAM30-Ssc matrix from A*0301 to B*2705 (Ssc_{ModB}). This matrix had a higher AUC (0.81) than the original matrix (Fig. 5). This result showed that predictions could potentially be improved by selecting best human matches based on similarity in terms of binding preferences if binding information is available. However, it is important to note that this improvement in the AUC was specific to the set of peptides available to date, and further prospective studies are needed to validate the preference of this particular pocket.

Overall, these results demonstrate that matrices built using contact residues from SLA structures and using the PAM30 substitution matrix to identify the best human match for each pocket, had the best predictive power of the approaches that were tested. Although the matrices showed predictive power, the limited number of known binders makes the AUC values less robust. For this reason, an analysis of a larger dataset of SLA-specific binders and non-binder peptides will be required to revalidate the predictive power of the matrices.

Discussion and conclusions

Immunoinformatics tools have accelerated the identification of epitopes and design of human vaccines. However, comparable tools have not been applied extensively to pigs. During the last two decades,

swine T cell epitope discovery has been based on experimental studies of numerous overlapping peptides [37–39, 48–54]. While these studies are essential for validating T cell epitope prediction tools, they can be expensive and time consuming. To reduce experimental effort and expedite the process, algorithms developed for human T cell epitope prediction have also been used to identify porcine epitopes [55–57]. However, the substitution of HLA predictions for SLA predictions may not reflect the fine specificity of SLA binding, which limits the efficacy of this oversimplified approach. To overcome this, we have developed PigMatrix, a simple yet effective method that leverages available data (SLA-binding peptides, SLA structures and HLA binding data) and pocket profiles already constructed for HLA-based epitope prediction in EpiMatrix to predict potential T cell epitopes for SLA class I and II alleles. Using the pocket profile method and the concept that pockets that have similar amino acids will share similar binding preferences, we built and validated matrices that were able to separate SLA-restricted peptides from random peptides and non-binders.

Human pan-specific tools based on the pocket profile method have been described for prediction of class I (PickPocket) and class II T cell epitopes (TEPITOPEpan) [9, 18]. These methods defined the amino acids in the pockets from HLA crystal structures. Similarly, we defined SLA-specific contacts (Ssc) from two crystal structures available for SLA-1*0401. Additionally, we extrapolated the pocket residues from crystal structures of HLA class I and II crystal structures (Hc). Both approaches assume for a given pocket that all contact residues are conserved across all class-specific MHC molecules. However, because there are differences in the pocket residues between MHC alleles and even between the same allele structures depending on the ligand [2], this simplification is a limitation of the peptide:MHC modeling approach. Even so, it is a reasonable approximation when structural information is limited. For class I, differences in the pocket residues using the Hc and the Ssc approaches were noticeable and impacted the subsequent selection of the best human match to build the prediction matrices. Matrices based on pockets defined from SLA structure-specific contacts performed better than HLA-derived pockets. While a definition of the contact residues based on several HLA structures account for the intra- and inter-allelic variability of binding pockets, it also dilutes the importance of key residues in the peptide:MHC interaction. We speculate that more allele-specific pockets could potentially improve the selection of the best human match and therefore the predictions.

Selection might be also improved by weighting the similarity score by conservation of key contact residues.

PickPocket and TEPITOPEpan use a method based on BLOSUM62 to calculate a weighted score of specificity to define the most similar HLA-derived pocket. For PigMatrix, in addition to BLOSUM62, we used PAM120, which is considered equivalent to BLOSUM62 for comparison of distantly related proteins [46], to calculate pocket similarity. We also included PAM30 and BLOSUM90, which are both designed to score similarity between closely related protein sequences. The SLA matrices with the highest AUC were based on PAM30 using both Hc and Ssc, with only one exception. If we consider the pocket contact residues as short pseudo-sequences, the better performance of PAM30-based matrices might be explained because low-numbered PAM matrices are more efficient for searches involving short sequences. BLOSUM62 on the other hand, performs better to identify distant homologs using longer sequences. BLOSUM90, like PAM30, is used for closely related sequences; however, it is not recommended for short peptides [46] and unsurprisingly did not perform as well as PAM30 in these studies.

Predictive methods for porcine T-cell epitopes are limited, and none existed previously for SLA class II. PigMatrix is the first tool to make binding predictions for an SLA-DR allele. Class II predictions were limited to the Hc method because no SLA-DR molecule has been crystallized. Since SLA-DR-specific binding data are scarce, predictions require further prospective validation. While it is not possible at this time to benchmark the SLA-DR matrix against other predictors, a comparison can be made for the SLA class I matrices developed here. NetMHCpan has been used for SLA binding predictions. PickPocket, which is also based on the pocket profile method, has been described primarily for HLA class I [18], but predictions are also available for SLA alleles. In this study, for an existing set of published peptides, PigMatrix performed equally or better than two versions of NetMHCpan and PickPocket for SLA-1*0401 and 2*0401. While PigMatrix and PickPocket derive SLA binding preferences from HLA binding data, NetMHCpan artificial neural networks are trained using information derived from available binding data as well as peptide sequences and MHC sequence information [6]. It was previously demonstrated that in a scenario where the quantitative binding data were limited for human and non-human MHC alleles, PickPocket performed better than NetMHCpan

[18]. This is also evident when NetMHCpan 2.4 results are compared to NetMHCpan 2.8 predictions. NetMHCpan 2.4 was trained with a limited set of SLA binders and its predictions were equivalent to random selection (average AUC 0.47). Version 2.8, on the other hand, was trained with more data and its performance improved for the alleles we evaluated (average AUC 0.76). Conversely, PigMatrix was not trained with SLA-specific binding data and performed similarly or better than NetMHCpan 2.8 predictions for two of three class I alleles we tested. Moreover, because the number of published peptides is limited, we were not able to compile a test dataset of peptides known to be different from the training set used by NetMHCpan 2.8. Hence, it is possible that NetMHCpan 2.8 performance was overestimated.

For SLA-3*0401, PickPocket and NetMHCpan 2.8 outperformed PigMatrix. Upon closer analysis, these results provided an example of how PigMatrix could be improved. We were able to build a better performing model by modifying pocket B in the matrix constructed using PAM30-Ssc. This might be explained by the role of the amino acid in position two of the peptide as a binding anchor and its specific interaction with this pocket. It is also worth noting that the HLA pocket library we used was limited to 12 class I and eight class II supertypes alleles available in EpiMatrix. It is possible that pocket sequences from other HLA alleles and their profiles are more similar to SLA pockets. Therefore, if the number of HLA alleles in the library is increased, we might find better human matches for SLA pockets, which could potentially improve matrix performance.

So as to illustrate the PigMatrix approach, we built initially matrices for only three SLA class I alleles and one SLA class II allele for which quantitative peptide data were available. These alleles are commonly expressed in different porcine breeds and cell lines for in vitro culture [32–36]. However, like HLA, SLA diversity is considerable. These results demonstrate the potential of the approach to be extended to SLA alleles with limited or nonexistent epitope binding data. Thus, future versions of PigMatrix will include a more comprehensive and representative set of matrices for SLA alleles expressed in outbred porcine populations. Moreover, prospective in vitro and in vivo evaluation of PigMatrix predictions will help to refine the matrices.

We developed the PigMatrix tool with the intent to integrate it into the iVAX toolkit, which is a comprehensive set of tools for computational vaccine design that includes EpiMatrix, Conservatrix,

ClustiMer, EpiAssembler, JanusMatrix, and VaccineCAD [58]. When the PigMatrices are used to substitute for HLA matrices (EpiMatrix) in iVAX, all of the existing suite of iVAX vaccine design tools can be used with the SLA epitope predictions, which makes it possible to envision accelerated development of novel T cell epitope-based vaccines or whole subunit vaccines optimized for epitope content that protect against infectious disease in swine.

Competing interests

AHG declares no competing interests. ADG and WDM are senior officers and majority shareholders at EpiVax, Inc., a privately-owned immunoinformatics and vaccine design company located in Providence, RI USA. LM and FT are employees at EpiVax, in which LM holds stock options. These authors acknowledge that there is a potential conflict of interest related to their relationship with EpiVax and attest that the work contained in this research report is free of any bias that might be associated with the commercial goals of the company. In addition to his role as a faculty member at Dartmouth, CBK is co-founder and CTO of Stealth Biologics, LLC, a therapeutic protein design company. Dartmouth has worked with him to manage all potential conflicts of interest arising from his commercial affiliation, and he affirms that this paper presents work free of any bias.

Author's contributions

AHG participated in the design of the study, development of the method, built the matrices, collected and analyzed the data, and prepared the manuscript. WDM transferred knowledge related to previous SLA development efforts and methods, provided consulting support in program design, organization of manuscript and figures. CBK participated in the design of the study and conception of the figures. FT provided technical support and edited the manuscript. LM gave conceptual advice and prepared the manuscript. ADG conceived of the project, reviewed the approach that was developed, provided advice and assisted with the development of the manuscript, contributed to, and edited the manuscript. All authors reviewed and approved the final manuscript.

Author's information

Anne S. De Groot MD is the director of the Institute for Immunology and Informatics (iCubed) in the College of the Environment and Life Sciences (CELS) at the University of Rhode Island (URI); Leonard Moise is the Director of the Vaccine and Immunotherapeutics Research Laboratory at the iCubed. The iCubed is dedicated to the promotion and development of computational vaccinology methods and tools for application to diseases of global relevance. Bill Martin is the architect of the iVAX suite of tools at EpiVax, and has generously provided access to those tools to the iCubed for animal and human vaccine research. Chris Bailey-Kellogg is Professor of Computer Science at Dartmouth, where his research focuses on computational structural biology and immunology. Andres H. Gutiérrez is a graduate student in the Department of Cell and Molecular Biology, CELS, URI.

Acknowledgments

The authors acknowledge the support of Jacob Tivin in the implementation of the matrices evaluated in this paper. Partial support for the development of PigMatrix was provided by the Institute for Immunology and Informatics and the National Pork Board (Project #12-121).

References

1. Neefjes J, Jongsma MLM, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat. Rev. Immunol. 2011;11:823–36.

2. Zhang N, Qi J, Feng S, Gao F, Liu J, Pan X, et al. Crystal structure of swine major histocompatibility complex class I SLA-1 0401 and identification of 2009 pandemic swine-origin influenza A H1N1 virus cytotoxic T lymphocyte epitope peptides. J. Virol. 2011;85:11709–24.

3. Wee LJK, Lim SJ, Ng LFP, Tong JC. Immunoinformatics: how in silico methods are re-shaping the investigation of peptide immune specificity. Front Biosci (Elite Ed). 2012;4:311–9.

4. Zhang Q, Wang P, Kim Y, Haste-Andersen P, Beaver J, Bourne PE, et al. Immune epitope database analysis resource (IEDB-AR). Nucleic Acids Res. 2008;36:W513–8.

5. Jojic N, Reyes-Gomez M, Heckerman D, Kadie C, Schueler-Furman O. Learning MHC I--peptide binding. Bioinformatics. 2006;22:e227–35.

6. Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, et al. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. PLoS ONE. 2007;2:e796.

7. Jacob L, Vert J-P. Efficient peptide-MHC-I binding prediction for alleles with few known binders. Bioinformatics. 2008;24:358–66.

8. Bordner AJ, Mittelmann HD. MultiRTA: a simple yet reliable method for predicting peptide binding affinities for multiple class II MHC allotypes. BMC Bioinformatics. 2010;11:482.

9. Zhang L, Chen Y, Wong H-S, Zhou S, Mamitsuka H, Zhu S. TEPITOPEpan: Extending TEPITOPE for Peptide Binding Prediction Covering over 700 HLA-DR Molecules. PLoS ONE. 2012;7:e30483.

10. Guo L, Luo C, Zhu S. MHC2SKpan: a novel kernel based approach for pan-specific MHC class II peptide binding prediction. BMC Genomics. 2013;14 Suppl 5:S11.

11. Karosiene E, Rasmussen M, Blicher T, Lund O, Buus S, Nielsen M. NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. Immunogenetics. 2013;65:711–24.

12. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. Immunogenetics. 2009;61:1–13.

13. Pedersen LE, Harndahl M, Rasmussen M, Lamberth K, Golde WT, Lund O, et al. Porcine major histocompatibility complex (MHC) class I molecules and analysis of their peptide-binding specificities. Immunogenetics. 2011;63:821–34.

14. Pedersen LE, Jungersen G, Buus S, Golde WT. Analysis of Swine Leukocyte Antigen peptide binding profiles and the identification of T cell epitopes by tetramer staining [Ph.D. thesis]. Technical University of Denmark; 2012.

15. Pedersen LE, Harndahl M, Nielsen M, Patch JR, Jungersen G, Buus S, et al. Identification of peptides from foot-and-mouth disease virus structural proteins bound by class I swine leukocyte antigen (SLA) alleles, SLA-1*0401 and SLA-2*0401. Anim. Genet. 2013;44:251–8.

16. Pedersen LE, Breum SO, Riber U, Larsen LE, Jungersen G. Identification of swine influenza virus epitopes and analysis of multiple specificities expressed by cytotoxic T cell subsets. Virol. J. 2014;11:163.

17. Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U, et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat. Biotechnol. 1999;17:555–61.

18. Zhang H, Lund O, Nielsen M. The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. Bioinformatics. 2009;25:1293–9.

19. Chelvanayagam G. A roadmap for HLA-A, HLA-B, and HLA-C peptide binding specificities. Immunogenetics. 1996;45:15–26.

20. Chelvanayagam G. A Roadmap for HLA-DR Peptide Binding Specificities. Human Immunology. 1997;58:61–9.

21. Rammensee H-G, Bachmann J, Stevanovic S. MHC Ligands and Peptide Motifs. Landes Bioscience, Georgetown, TX; 1997.

22. De Groot AS, Jesdale BM, Szu E, Schafer JR, Chicz RM, Deocampo G. An interactive Web site providing major histocompatibility ligand predictions: application to HIV research. AIDS Res. Hum. Retroviruses. 1997;13:529–31.

23. McMurry JA, Kimball S, Lee JH, Rivera D, Martin W, Weiner DB, et al. Epitope-driven TB vaccine development: a streamlined approach using immuno-informatics, ELISpot assays, and HLA transgenic mice. Curr. Mol. Med. 2007;7:351–68.

24. Gregory SH, Mott S, Phung J, Lee J, Moise L, McMurry JA, et al. Epitope-based vaccination against pneumonic tularemia. Vaccine. 2009;27:5299–306.

25. Moise L, Buller RM, Schriewer J, Lee J, Frey SE, Weiner DB, et al. VennVax, a DNA-prime, peptide-boost multi-T-cell epitope poxvirus vaccine, induces protective immunity against vaccinia infection by T cell response alone. Vaccine. 2011;29:501–11.

26. Moss SF, Moise L, Lee DS, Kim W, Zhang S, Lee J, et al. HelicoVax: epitope-based therapeutic Helicobacter pylori vaccination in a mouse model. Vaccine. 2011;29:2085–91.

27. De Groot AS, Martin W. Reducing risk, improving outcomes: bioengineering less immunogenic protein therapeutics. Clin. Immunol. 2009;131:189–201.

28. De Groot AS, Jesdale B, Martin W, Saint Aubin C, Sbai H, Bosma A, et al. Mapping cross-clade HIV-1 vaccine epitopes using a bioinformatics approach. Vaccine. 2003;21:4486–504.

29. Moise L, McMurry JA, Buus S, Frey S, Martin WD, De Groot AS. In silico-accelerated identification of conserved and immunogenic variola/vaccinia T-cell epitopes. Vaccine. 2009;27:6471–9.

30. Mishra S, Losikoff PT, Self AA, Terry F, Ardito MT, Tassone R, et al. Peptide-pulsed dendritic cells induce the hepatitis C viral epitope-specific responses of naïve human T cells. Vaccine. 2014;32:3285–92.

31. De Groot AS, Nene V, Hegde NR, Srikumaran S, Rayner J, Martin W. T cell epitope identification for bovine vaccines: an epitope mapping method for BoLA A-11. Int. J. Parasitol. 2003;33:641–53.

32. Smith DM, Lunney JK, Martens GW, Ando A, Lee J-H, Ho C-S, et al. Nomenclature for factors of the SLA class-I system, 2004. Tissue Antigens. 2005;65:136–49.

33. Smith DM, Lunney JK, Ho C-S, Martens GW, Ando A, Lee J-H, et al. Nomenclature for factors of the swine leukocyte antigen class II system, 2005. Tissue Antigens. 2005;66:623–39.

34. Ho C-S, Franzo-Romain MH, Lee YJ, Lee JH, Smith DM. Sequence-based characterization of swine leucocyte antigen alleles in commercially available porcine cell lines. Int. J. Immunogenet. 2009;36:231–4.

35. Ho C-S, Lunney JK, Franzo-Romain MH, Martens GW, Lee Y-J, Lee J-H, et al. Molecular characterization of swine leucocyte antigen class I genes in outbred pig populations. Anim. Genet. 2009;40:468–78.

36. Ho C-S, Lunney JK, Lee J-H, Franzo-Romain MH, Martens GW, Rowland RRR, et al. Molecular characterization of swine leucocyte antigen class II genes in outbred pig populations. Anim. Genet. 2010;41:428–32.

37. Ober BT, Summerfield A, Mattlinger C, Wiesmüller KH, Jung G, Pfaff E, et al. Vaccine-induced, pseudorabies virus-specific, extrathymic CD4+CD8+ memory T-helper cells in swine. J. Virol. 1998;72:4866–73.

38. Armengol E, Wiesmüller K-H, Wienhold D, Büttner M, Pfaff E, Jung G, et al. Identification of T-cell epitopes in the structural and non-structural proteins of classical swine fever virus. J. Gen. Virol. 2002;83:551–60.

39. Gerner W, Denyer MS, Takamatsu H-H, Wileman TE, Wiesmüller K-H, Pfaff E, et al. Identification of novel foot-and-mouth disease virus specific T-cell epitopes in c/c and d/d haplotype miniature swine. Virus Res. 2006;121:223–8.

40. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003;31:3784–8.

41. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. BMC Immunology. 2008;9:1.

42. Southwood S, Sidney J, Kondo A, del Guercio MF, Appella E, Hoffman S, et al. Several common HLA-DR types share largely overlapping peptide binding repertoires. J. Immunol. 1998;160:3363–73.

43. Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, et al. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. Nature. 1994;368:215–21.

44. Stern LJ, Calvo-Calle JM. HLA-DR: molecular insights and vaccine design. Curr. Pharm. Des. 2009;15:3249–61.

45. Dayhoff MO, Schwartz RM. A model of evolutionary change in proteins. Atlas of Protein Sequence and Structure. 1978.

46. Altschul SF. Amino acid substitution matrices from an information theoretic perspective. J. Mol. Biol. 1991;219:555–65.

47. Madden DR. The three-dimensional structure of peptide-MHC complexes. Annu. Rev. Immunol. 1995;13:587–622.

48. Blanco E, McCullough K, Summerfield A, Fiorini J, Andreu D, Chiva C, et al. Interspecies major histocompatibility complex-restricted Th cell epitope on foot-and-mouth disease virus capsid protein VP4. J. Virol. 2000;74:4902–7.

49. Blanco E, Garcia-Briones M, Sanz-Parra A, Gomes P, De Oliveira E, Valero ML, et al. Identification of T-cell epitopes in nonstructural proteins of foot-and-mouth disease virus. J. Virol. 2001;75:3164–74.

50. García-Briones MM, Blanco E, Chiva C, Andreu D, Ley V, Sobrino F. Immunogenicity and T cell recognition in swine of foot-and-mouth disease virus polymerase 3D. Virology. 2004;322:264–75.

51. Stevenson LS, Gilpin DF, Douglas A, McNeilly F, McNair I, Adair BM, et al. T lymphocyte epitope mapping of porcine circovirus type 2. Viral Immunol. 2007;20:389–98.

52. Vashisht K, Goldberg TL, Husmann RJ, Schnitzlein W, Zuckermann FA. Identification of immunodominant T-cell epitopes present in glycoprotein 5 of the North American genotype of porcine reproductive and respiratory syndrome virus. Vaccine. 2008;26:4747–53.

53. Wang Y-X, Zhou Y-J, Li G-X, Zhang S-R, Jiang Y-F, Xu A-T, et al. Identification of immunodominant T-cell epitopes in membrane protein of highly pathogenic porcine reproductive and respiratory syndrome virus. Virus Res. 2011;158:108–15.

54. Parida R, Choi I-S, Peterson DA, Pattnaik AK, Laegreid W, Zuckermann FA, et al. Location of T-cell epitopes in nonstructural proteins 9 and 10 of type-II porcine reproductive and respiratory syndrome virus. Virus Res. 2012;169:13–21.

55. Díaz I, Pujols J, Ganges L, Gimeno M, Darwich L, Domingo M, et al. In silico prediction and ex vivo evaluation of potential T-cell epitopes in glycoproteins 4 and 5 and nucleocapsid protein of genotype-I (European) of porcine reproductive and respiratory syndrome virus. Vaccine. 2009;27:5603–11.

56. Zimic M, Gutiérrez AH, Gilman RH, López C, Quiliano M, Evangelista W, et al. Immunoinformatics prediction of linear epitopes from Taenia solium TSOL18. Bioinformation. 2011;6:271–4.

57. Burgara-Estrella A, Díaz I, Rodríguez-Gómez IM, Essler SE, Hernández J, Mateu E. Predicted peptides from non-structural proteins of porcine reproductive and respiratory syndrome virus are able to induce IFN-γ and IL-10. Viruses. 2013;5:663–77.

58. Terry FE, Moise L, Martin RF, Torres M, Pilotte N, Williams SA, et al. Time for T? Immunoinformatics addresses vaccine design for neglected tropical and emerging infectious diseases. Expert Rev Vaccines. 2015;14:21–35.

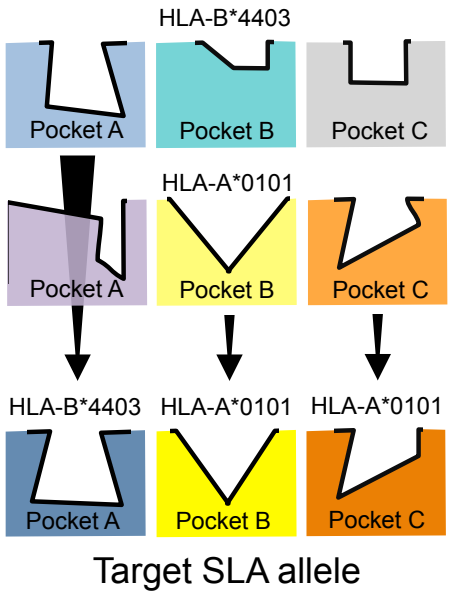


Fig. 1. Illustration of PigMatrix development using the pocket profile method. Three pockets (A, B, and C) from human (HLA) and swine (SLA) MHC molecules are represented as different shapes and colors. The contours of the pockets are shown in bold black lines. HLA pockets from two HLA alleles (HLA-A*0101 and B*4403) are shown in the first two rows. For each pocket in a target SLA, in the third row, we identified the most similar HLA pocket (best human match) and combined their pocket profiles (binding preferences expressed as coefficients) to build composite predictive matrices (PigMatrix).

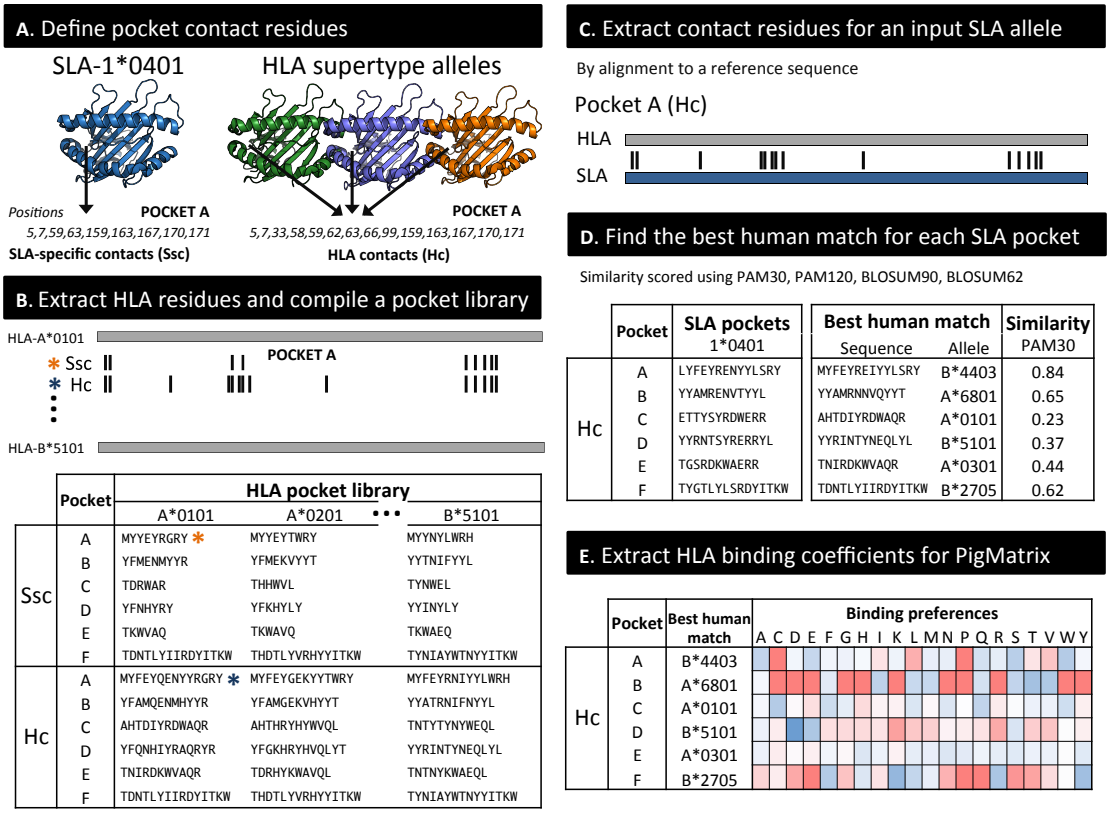
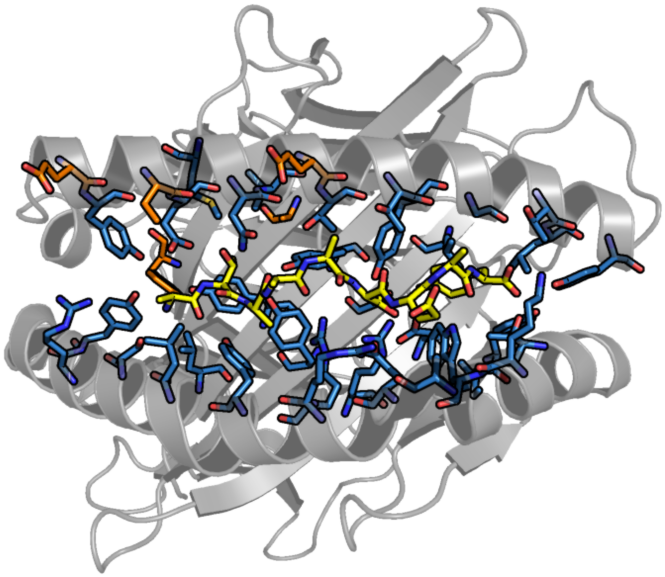


Fig. 2. PigMatrix algorithm. (A) Residues in contact with the peptide are determined with respect to a crystal structure of either an SLA allele (Ssc) or an HLA supertype allele (Hc). Class I supertype alleles are represented by three HLA molecules. (B) Contact residues defined by either Ssc or Hc are extracted from HLA sequences and compiled into a library of HLA pockets (HLA pocket library). Pocket A positions and the extracted pocket sequences for Ssc and Hc are marked with *. (C) For a target SLA allele, contact residues (defined by Ssc or Hc) are identified by aligning the SLA sequence to a reference HLA sequence. (D) SLA pocket sequences are compared to those in the HLA pocket library to identify the best human match. (E) Binding coefficients of the best human match for each SLA pocket sequence are compiled to build a PigMatrix. Coefficients are represented in red to blue scale (high to low binding likelihood).



SLA position	SLA-specific contacts (Ssc) Pocket (peptide position) A(1)	B(2)	C(6)	D(3)	E(7)	F(9)	Count	HLA contacts (Hc) Pocket (peptide position) A(1)	B(2)	C(6)	D(3)	E(7)	F(9)	Count
5L	■						1	■						1
7Y	■			■			3	■	■		■			3
9Y		■		■			2		■		■			2
24A							0		■					1
33F							0	■						1
45M		■					1		■					1
58E							0	■						1
59Y	■						1	■						1
62R							0	■	■		■			3
63E	■	■					2	■	■					2
66N		■		■			2	■	■		■			3
67V		■					1		■					1
69E							0			■				1
70T				■			1		■	■				3
73T			■		■	■	3			■		■	■	3
74Y			■			■	2			■			■	2
77G						■	1					■	■	2
80T						■	1						■	1
81L						■	1						■	1
84Y						■	1						■	1
95L						■	1						■	1
97S						■	1			■			■	4
99Y		■		■			2	■	■	■				4
114R			■			■	2			■			■	4
116D						■	1			■		■	■	3
123Y						■	1						■	1
124I						■	1						■	1
143T						■	1						■	1
146K					■	■	2					■	■	2
147W			■		■	■	3			■		■	■	3
150A					■		1					■		1
152E			■		■		2			■	■	■		3
155R					■		1			■	■	■		3
156R			■	■			2			■	■	■		3
159Y	■			■			3	■			■			3
163L	■	■					2	■	■		■			3
167S	■						1	■						2
170R	■						1	■						1
171Y	■						1	■						1
Total	9	9	6	7	6	15		14	12	12	13	11	15	

Fig. 3. Comparison of contact residues in the binding pockets based on SLA-specific contacts (Ssc) and HLA contacts (Hc). Top: Schematic representation of the crystal structure of SLA-1*0401 (PDB:3QQ4; residues 1 to 181 rendered with PyMOL (Schrodinger, LLC)) showing the residues involved in the binding pockets. SLA contact residues and the ligand (ATAAATEAY, yellow) are represented as sticks. Residues common for both Hc and Ssc approaches are show in light blue; residues unique to Hc in orange. **Bottom:** Positions in the SLA binding pockets are shown. The first column (SLA position) is the residue and position number in the SLA-1*0401 protein sequence (Genbank:2352988). Residue positions shown in bold and underlined are identical (i.e. amino acid involved in the same pocket(s)) for both approaches. Positions in light blue are common for both approaches; positions in orange are unique to Hc. The next columns show, shaded in gray, the positions involved in pockets A through F that interact with relative ligand positions (peptide position). The last column (Count) is the number of pockets in which an amino acid participates. The last row (Total) is the total number of residues in each pocket.

SLA allele	Similarity score	Mean binding likelihood expressed as Z-score (sd)			
		Ssc		Hc	
		Binders	Non-binders	Binders	Non-binders
1*0401	PAM30	**2.12 (0.46)**	**1.49 (0.57)**	1.65 (0.55)	1.79 (0.48)
	BLOSUM90	2.98 (1.00)	3.06 (1.00)	2.99 (0.96)	3.15 (0.97)
	PAM120	**2.12 (0.46)**	**1.49 (0.57)**	1.61 (0.54)	1.69 (0.51)
	BLOSUM62	2.98 (1.00)	3.06 (1.00)	3.02 (0.97)	3.17 (0.96)
	Negative	-0.58 (0.76)	-0.45 (0.78)	0.51 (0.67)	0.53 (0.52)
2*0401	PAM30	**2.30 (0.76)**	**1.45 (0.84)**	2.07 (0.83)	1.83 (1.10)
	BLOSUM90	**2.30 (0.76)**	**1.45 (0.84)**	2.07 (0.83)	1.83 (1.10)
	PAM120	**2.30 (0.76)**	**1.45 (0.84)**	2.07 (0.83)	1.83 (1.10)
	BLOSUM62	**2.30 (0.76)**	**1.45 (0.84)**	2.20 (0.98)	1.96 (1.09)
	Negative	**0.56 (0.59)**	**1.00 (0.75)**	**0.56 (0.59)**	**1.00 (0.75)**
3*0401	PAM30	1.42 (1.05)	1.16 (0.77)	1.34 (1.17)	1.12 (0.78)
	BLOSUM90	**1.62 (0.81)**	**2.20 (0.82)**	1.71 (0.90)	1.96 (0.82)
	PAM120	**1.36 (0.72)**	**1.95 (0.78)**	**1.27 (1.06)**	**1.68 (0.95)**
	BLOSUM62	**1.74 (0.79)**	**2.22 (0.75)**	1.98 (0.90)	2.32 (0.87)
	Negative	0.52 (0.70)	0.47 (0.55)	**0.54 (0.82)**	**1.13 (0.96)**
DRB1*0201	PAM30			**1.85 (0.65)**	**1.29 (0.52)**
	BLOSUM90			**1.67 (0.57)**	**1.36 (0.55)**
	PAM120			**1.67 (0.57)**	**1.36 (0.55)**
	BLOSUM62			**1.67 (0.57)**	**1.36 (0.55)**
	Negative			**0.30 (0.42)**	**-0.03 (0.43)**

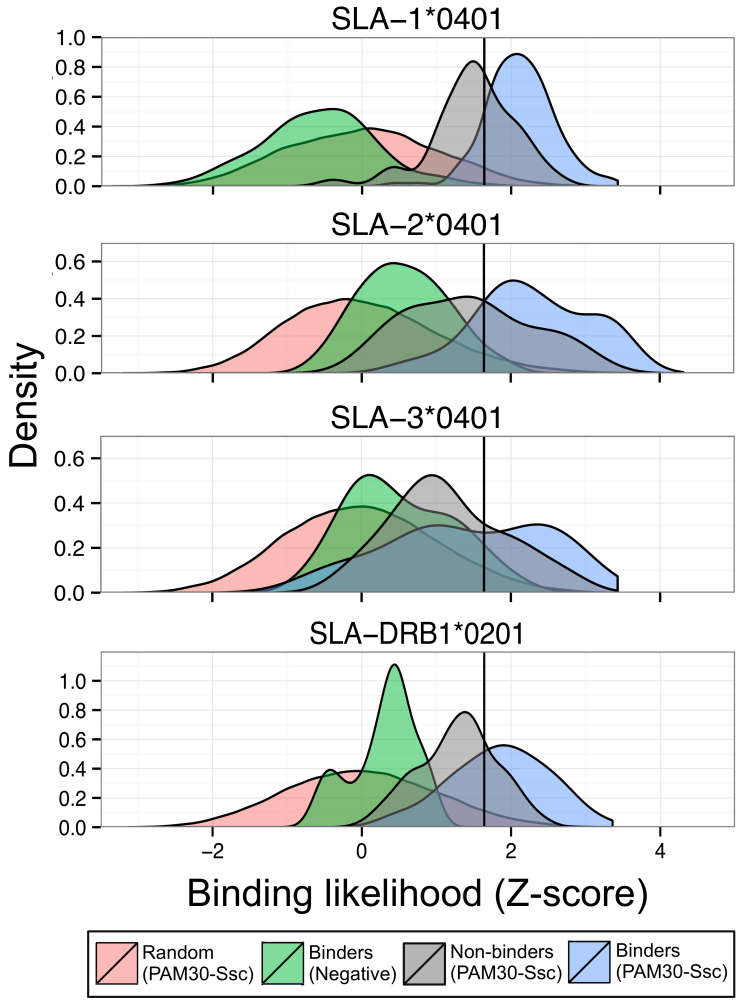


Fig. 4. Validation of composite matrices. Top: Binding likelihood (Z-score) means and standard deviations (sd) of binders and non-binders calculated using the matrices built based on Ssc and Hc and different scoring methods for pocket selection are shown for SLA class I and II alleles. Z-score means and sd calculated using negative matrices, HLA matrices with the lowest overall pocket identity for each SLA allele, are also shown. Instances where the Z-scores of binders and non-binders were statistically different (p-value ≤ 0.05) using a Wilcoxon-Mann–Whitney test, are shown in bold; those with Z-score means above or below 1.64 for binders or non-binders, respectively, are shown in gray. **Bottom:** Comparison of binding likelihood (expressed as Z-score) between matrices (PAM30-Ssc and Negative) shown as density estimates (smoothed histograms). Note that y-axes are differently scaled. Binders and non-binders were scored with PAM30-Ssc (for class I), PAM30-Hc (for class II) and Negative control matrices. 100,000 natural random 9-mers were scored with either PAM30-Ssc (class I) or PAM30-Hc (class II). The black line indicates the threshold at which a 9-mer is considered a potential binder (Z-score of 1.64). Ssc was not applied to SLA-DRB1*0201 because crystal structures are not available.

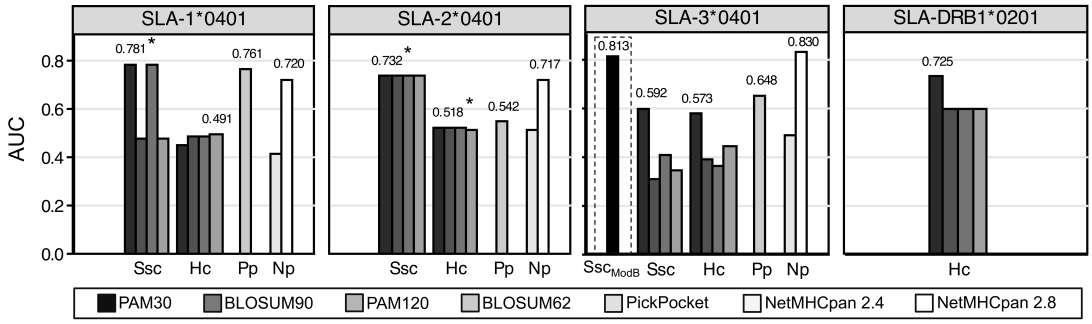


Fig. 5. Matrix performance comparison. AUCs of the matrices built for SLA class I and II alleles are shown. The highest AUC for each method is shown above the bars. For SLA-3*0401, AUC of Ssc_{ModB}, a PAM30-Ssc matrix with a pocket B profile different than the best human match, is shown in a dashed rectangle to illustrate the impact of pocket B. If one or more matrices for the same approach (Ssc or Hc) have equal AUC, it is indicated with *.

CHAPTER 3

Manuscript published in PLOS ONE, July 2016

In vivo **validation of predicted and conserved T cell epitopes in a swine influenza model**

Andres H. Gutiérrez[1], Crystal Loving[2], Leonard Moise[1,3], Frances E. Terry[3], Susan L. Brockmeier[2], Holly R. Hughes[2], William D. Martin[3], Anne S. De Groot[1,3]*

[1]Institute for Immunology and Informatics, Department of Cell and Molecular Biology, University of Rhode Island, Providence, RI, USA

[2]Virus and Prion Diseases Research Unit, NADC, USDA ARS, Ames, IA, USA

[3]EpiVax Inc., Providence, RI, USA

*Corresponding Author: Anne S. De Groot, M.D.

Institute for Immunology and Informatics

Department of Cell and Molecular Biology

University of Rhode Island

Providence, RI, 02903, USA

E-mail: dr.annie.degroot@gmail.com

Abstract

Swine influenza is a highly contagious respiratory viral infection in pigs that is responsible for significant financial losses to pig farmers annually. Current measures to protect herds from infection include: inactivated whole-virus vaccines, subunit vaccines, and alpha replicon-based vaccines. As is true for influenza vaccines for humans, these strategies do not provide broad protection against the diverse strains of influenza A virus (IAV) currently circulating in U.S. swine. Improved approaches to developing swine influenza vaccines are needed. Here, we used immunoinformatics tools to identify class I and II T cell epitopes highly conserved in seven representative strains of IAV in U.S. swine and predicted to bind to Swine Leukocyte Antigen (SLA) alleles prevalent in commercial swine. Epitope-specific interferon-gamma (IFNγ) recall responses to pooled peptides and whole virus were detected in pigs immunized with multi-epitope plasmid DNA vaccines encoding strings of class I and II putative epitopes. In a retrospective analysis of the IFNγ responses to individual peptides compared to predictions specific to the SLA alleles of cohort pigs, we evaluated the predictive performance of PigMatrix and demonstrated its ability to distinguish non-immunogenic from immunogenic peptides and to identify promiscuous class II epitopes. Overall, this study confirms the capacity of PigMatrix to predict immunogenic T cell epitopes and demonstrate its potential for use in the design of epitope-driven vaccines for swine. Additional studies that match the SLA haplotype of animals with the study epitopes will be required to evaluate the degree of immune protection conferred by epitope-driven DNA vaccines in pigs.

Keywords: PigMatrix, epitope prediction, T cell epitope, Immunogenicity, Swine, influenza, influenza A virus, vaccine, immunoinformatics, Swine Leukocyte Antigen, SLA, iVAX, class I, class II

Introduction

Swine influenza is a highly contagious respiratory viral infection in pigs that has a major impact on their health. In addition, influenza outbreaks are responsible for significant financial losses to pig farmers, large and small, on an annual basis [1]. The negative economic impact is due to weight loss, reduced weight gain and predisposition to other infections [2]. Clinical signs of the disease include fever, coughing, sneezing, nasal discharge, lethargy, and anorexia. The causative agent is influenza A virus (IAV), a negative-sense, single-stranded, segmented RNA virus of the *Orthomyxoviridae* family. Transmission is by direct contact and by aerosol [3]. As is true with IAV in humans, antigenic drift by accumulation of mutations and/or antigenic shift by reassortment with genes from other IAV subtypes results in the emergence of novel influenza viruses [4]. Human-to-swine 'spillover' events also contribute to the genetic diversity of swine IAV [5]. H1N1, H1N2, and H3N2 swine IAV subtypes are endemic and co-circulate in swine in the U.S. [6].

Continual reassortment events led to the emergence of a novel triple-reassortant internal gene (TRIG) cassette that contains internal genes derived from human (PB1 gene), avian (PA and PB2 genes) and swine (NS, NP, and M genes) IAV viruses [7]. The TRIG is conserved among swine IAV circulating subtypes and it seems to have the ability to combine with numerous hemagglutinin (HA) and neuraminidase (NA) genes, including those of human and swine origin leading to enhanced strain variability [7]. Thus, the primary antigenic component of swine IAV vaccines is HA, which has evolved to present antigenically distinct HA lineages including: (1) the classical swine lineages, H1α, H1β, H1γ, H1γ-2, H1pdm09; (2) lineages derived from human seasonal H1 viruses, H1δ1, H1δ2; and (3) H3 cluster I-IV viruses [6,8,9]. This marked genetic diversity complicates the development of effective vaccines for pigs.

The predominant type of vaccine used by pork producers consists of whole inactivated viruses (WIV), administered with adjuvant by intramuscular injection. HA is the primary target of protective antibody responses of this platform. These vaccines are problematic for three reasons. First, antibody induced by WIV vaccination does not provide significant protection against antigenically diverse strains of IAV [8,10]. Second, WIV vaccines have been linked to vaccine-associated enhanced respiratory disease (VAERD) in pigs when WIV vaccine and infecting strains are mismatched [11–13]. Lastly,

existing vaccines do not adequately address viral diversity.

In contrast, cell-mediated immune responses to epitopes that are conserved across IAV strains have been shown, in a number of studies, to be protective against influenza. For example, human and mouse studies demonstrate that cell-mediated responses to conserved non-structural proteins can be broadly cross-reactive [14] and protective against variety of IAV subtypes [15]. Both CD4+ T helper cells (Th) and CD8+ cytotoxic T cells (CTL) contribute to clearance of IAV [16–18]. T cell help is also required for the development of high titers of strain-specific antibody [18]. In fact, memory T cell response improves vaccine efficacy against emerging IAV strains when cross-reactive helper T cell populations are present from prior infection and/or vaccination [19]. CTL responses have also been associated with viral clearance and reduced clinical severity in mice and humans [20,21]. Our group has been interested in the role of cross-conserved epitopes in protection against IAV in human populations, and has postulated that immunity to cross-conserved epitopes may have contributed to attenuation of morbidity in some age groups during the 2009 H1N1 IAV pandemic [22].

Adaptive cell-mediated immune response depends on T cell receptor (TCR) recognition of peptides bound to major histocompatibility complex (MHC) molecules presented on the surface of cells. Immunoinformatics tools have accelerated the discovery of T cell epitope peptides and design of epitope-driven vaccines (EDV) for human IAV [23–26]. The lack of quantitative MHC binding data has limited the development of tools for swine, cattle, and other food animal species. We recently developed a new tool for swine epitope prediction (PigMatrix) that leverages the pocket profile method originally described by Sturniolo et al. [27]. We integrated the new swine MHC predictions into iVAX, the suite of tools for vaccine design that were validated in a number of pre-clinical studies of human vaccines [28,29]. This set of tools is particularly useful for identifying T cell epitopes that are conserved across subtypes of strains [30], which is relevant to develop a IAV vaccine for pigs. Having integrated the new matrices into this 'in silico vaccine design' platform, we were able to apply the PigMatrix version of iVAX to IAV.

In this study, we used PigMatrix to predict class I and II T cell epitopes that are conserved in external and internal proteins from seven circulating IAV strains. We selected epitopes predicted to bind to SLA alleles that were previously reported to be prevalent in outbred U.S. swine populations

[31,32] and developed a prototype PigMatrix epitope-driven DNA-vaccine (PigMatrix-EDV) as a tool to evaluate immunogenic responses to highly conserved predicted epitopes in a swine IAV model. PigMatrix predicted peptides induced specific interferon gamma (IFNγ) recall responses in pigs immunized with the prototype PigMatrix-EDV encoding strings of class I and II putative epitopes. In addition, we performed a retrospective analysis to compare IFNγ responses to individual peptides (28 class I and 20 class II peptides) with predictions specific to the SLA expressed in the study cohort. The results showed that cohort-specific predictions using PigMatrix, were particularly effective for identification of non-immunogenic peptides.

Materials and Methods

Sequences

Gene sequences of proteins expressed by seven representative swine IAV (pandemic A/California/04/2009 (H1N1) (H1N1pdm09), A/swine/Illinois/5265/2010 (H1N1) (IL/10), A/swine/Ohio/511445/2007 (H1N1) (OH/07), A/swine/Minnesota/02011/2008 (H1N2) (MN/08), A/swine/Minnesota/A01301731/2012 (H1N2) (MN/12), A/swine/Texas/4199-2/1998 (H3N2) (TX/98), A/turkey/Ohio/313053/2004 (H3N2) (OH/04)) [33,34] were downloaded from the Influenza Virus Resource [35] (Supplemental Table 1).

Conservation analysis

The goal of the conservation analysis was to identify highly cross-conserved 9-mer peptides. Since 9-mers fit into the SLA binding groove [36], proteins derived from IAV genomes were parsed into 9-mer frames overlapping by eight amino acids using the Conservatrix algorithm [30]. Nine-mer sequences were searched for identically matched segments among IAV strains, as previously described [28]. Resulting 9-mers were ranked by their conservation within the dataset.

T cell epitope prediction

Using the pocket profile method [27] and well-defined EpiMatrix binding preferences for human MHC pockets, we developed PigMatrix prediction matrices as previously described [29]. Matrices were designed based on the binding preferences of the best-matched Human Leukocyte Antigen (HLA) pocket for each SLA pocket. The contact residues involved in the binding pockets were defined from

crystal structures of SLA or HLA supertype alleles for class I and II, respectively. Allele selection was based on prior data indicating their prevalence in outbred swine populations [31,32]. Matrices were constructed to predict T cell epitope binding to class I (SLA-1*0101, 1*0401, 2*0101, and 2*0401) and class II (SLA-DRB1*0101, 0201, 0401, and 0601) SLA alleles. SLA-1*0401, 2*0401 and SLA-DRB1*0201 were previously validated using published epitopes [29]. We also developed matrices for SLA alleles expressed in the study cohort (cohort-specific prediction) to perform a retrospective analysis.

All highly conserved 9-mers resulting from Conservatrix analysis were scored for binding potential against the panel of SLA alleles. PigMatrix raw scores were standardized to Z-scores to compare potential epitopes across multiple SLA alleles. Peptides with Z-scores above 1.64 (the top 5% of any given sample of 9-mers) were identified as likely to be SLA ligands. The final selection of putative SLA class I-restricted epitopes was based on PigMatrix score (Z-score>1.64), SLA class I allele coverage (≥50%) and IAV strain coverage.

Construction of immunogenic consensus sequences

EpiAssembler was used to construct 16-25 amino acid length SLA-DRB1-restricted sequences that were highly conserved in IAV strains, promiscuous (predicted to bind to multiple alleles), and enriched for immunogenicity (immunogenic consensus sequences or ICS) [28]. The density of predicted binding motifs in each ICS was scored (i.e. cluster score) using ClustiMer [28]. The cluster score represents the deviation in predicted epitope content from baseline expectation based on random peptides [37]. ICS with cluster scores above 10 were considered to be high-quality clusters for inclusion in the prototype vaccine. Peptides were ranked based on cluster score and IAV strain coverage and the final selection of epitopes was made using the same three criteria described above for class I peptides. Highly hydrophobic peptides were excluded as these are known to be more technically difficult to synthesize and may be less soluble in aqueous solutions.

Multi-epitope plasmid DNA vaccine engineering and production

Predicted epitope sequences were concatenated to form two multi-epitope genes (one for SLA class I and one for class II epitopes). VaccineCAD [38] and a concatemer optimization algorithm (unpublished) were used to rearrange the peptides to avoid creation of novel epitopes at peptide

junctions and to search for transmembrane helices that might interfere with production of the epitope concatemer proteins. Both algorithms, VaccineCAD and the concatemer optimization algorithm, used PigMatrix to predict junctional epitopes.

Transmembrane helices were predicted using TMHMM 2.0 [39]. In addition, where reordering did not sufficiently reduce the potential for junctional immunogenicity, a cleavage promoting spacer ('AAY') for class I-restricted constructs [40] or a binding inhibiting 'breaker' sequence ('GPGPG') for class II-restricted constructs [41], was introduced between peptides to optimize epitope processing. Two genes (one for class I and one for class II epitopes) predicted to have no transmembrane segments or junctional epitopes, were codon-optimized and synthesized by GeneArt (Life Technologies, NY, USA). Tandem stop codons were incorporated downstream of the epitope sequences. Class I and class II genes, respectively, were subcloned at predefined flanking restriction sites downstream of either a destabilizing UbiquitinA76 tag (UbA76) in pNTC8684-eRNA41H for proteasome targeting and a tissue plasminogen activator (TPA) leader sequence in pNTC8682-eRNA41H (Nature Technology Corporation, NE, USA) for secretory pathway targeting. High-purity plasmids for immunizations were prepared by Nature Technology Corporation, Inc. at research grade. Each plasmid underwent quality control testing including spectrophotometric concentration and A_{260}/A_{280} ratio determination (1.97), restriction digest analysis to assure the presence of the multi-epitope genes, agarose gel electrophoresis determination of residual host RNA and DNA (none detected), and quantitative endotoxin testing (<2.0 EU/mg).

Peptide synthesis

Peptides corresponding to putative epitopes in the DNA vaccine were synthesized using 9-fluoronylmethoxycarbonyl (Fmoc) chemistry by 21st Century Biochemicals (Marlboro, MA). Peptide purity was >80% as ascertained by analytical reversed phase HPLC. Peptide mass was confirmed by tandem mass spectrometry.

Immunizations

Thirty-two, 3-week old outbred pigs from a high-health status herd known to be free of IAV were delivered to the USDA-National Animal Disease Center. To ensure that prior exposure to IAV resulting in immunity was absent, all of the pigs were screened for influenza A nucleoprotein antibody by ELISA

(MultiS ELISA, IDEXX, Westbrook, Maine) prior to the start of the study. All of the study pigs were treated with ceftiofur crystalline-free acid (Excede; Zoetis Animal Health, Florham Park, NJ, USA) and enrofloxacin (Baytril 100; Bayer HealthCare AG, Monheim, Germany) upon arrival to reduce bacterial contaminants. The experimental outline is summarized in Fig. 1. Pigs were randomly distributed into four groups of eight and housed in separate isolation rooms in animal biosafety level 2 (ABSL-2) containment. Three groups were vaccinated: (i) one group of eight pigs was vaccinated with the prototype PigMatrix DNA-vaccine as the initial prime vaccination, followed by two homologous boosts at 21 and 42 days post-initial vaccination (dpv) (PigMatrix-EDV); (ii) one group of eight pigs was vaccinated with empty DNA plasmids containing no epitopes (Sham); (iii) one group of eight pigs was vaccinated with commercially available FluSureXP® administered 21 days apart, according to the manufacturer's directions (Zoetis Animal Health, Florham Park, NJ) (FluSure). FluSureXP® contains whole inactivated γ-cluster H1N1, δ1-cluster H1N1, δ2-cluster H1N1, and cluster IV H3N2 swine IAV viruses. The final group of eight non-vaccinated pigs served as controls (NV). The prototype PigMatrix-EDV plasmids were thawed at 4°C overnight, combined and administered intramuscularly in the postauricular region of the neck by needle stick injection with 4 mg per DNA plasmid in 4 mL of Tris-EDTA (TE) buffer (2 mL on right side and 2 mL on left side).

Animal care

Animals at the National Animal Disease Center (NADC) are cared for in accordance with the guidelines set forth in the "Guide for the Care and Use of Laboratory Animals" (National Academy Press, 1996) and in regulations and standards as promulgated by the Agricultural Research Service, USDA, pursuant to the Laboratory Animal Welfare Act of August 24, 1966, as amended. Animal studies are reviewed and approved by NADC's Institutional Animal Care and Use Committee (IACUC). In addition, the IACUC is federally mandated to review, at least once every 6 months, the research facility's animal care program and physical facilities per USDA regulations and using the "Guide for the Care and Use of Laboratory Animals" as the basis for review. Full-time animal caretakers, technicians and supervisors and on-call veterinarians perform routine animal care, as well as weekend/holiday activities and respond to emergencies. NADC staff members who worked with the animals have backgrounds and continuing training in the appropriate, species-specific care and handling of research animals. Training

courses for animal staff include safe handling skills, animal welfare, specific procedures (e.g. bleeding), personal protective equipment (PPE) requirements, as well as proper handling and care and use of anesthetics and analgesics. For this study, animals were housed in an ABSL-2 facility (12 h light/dark cycle) during the course of the study, and humanely euthanized at the termination of the project with a lethal dose of pentobarbital (Fatal Plus; Vortech Pharmaceuticals, Dearborn, MI). Protocols were in place to humanely euthanize any animals if unforeseen clinical disease presented, such as severe lameness or depression that results in recumbency with reluctance to stand, although that did not occur in this study (all animals in the study were terminated at the end of the experiment). Animal observations and feedings were completed at least twice daily by personnel who have been trained to look for signs of illness or abnormalities, at which time the veterinarian on-call and the principal investigator would have been notified.

Measurement of IFNγ response by ELISpot assay

At 42, 49 and 63 dpv, whole blood was collected by venipuncture and peripheral blood mononuclear cells (PBMC) were isolated as previously described [42]. The frequency of epitope-specific T cells was determined by porcine IFNγ enzyme-linked immunosorbent spot (IFNγ ELISpot) assay according to the manufacturer's recommendations (R&D Systems, Minneapolis, MN). Wells were seeded with 2.5 x 10^5 PBMCs and stimulated with pooled peptides at 10 μg/mL, whole H1N1pdm09 virus (WV) at a multiplicity of infection (MOI) of 0.5, pokeweed mitogen (PWM) at 1 μg/mL, or culture media in a final volume of 0.25 mL.

Immune responses to the IAV epitopes contained in the vaccines were evaluated using PBMC from each of 32 study animals. To simplify the analysis, four pools of peptides were evaluated at all PBMC sampling points – one that included all 48 predicted peptides (All); a second pool that contained 26 class I and II peptides predicted from internal proteins (Int); a third pool that contained 8 class II peptides predicted from external proteins (Ext-II); and a final pool that contained 14 class I peptides predicted from internal proteins (Int-I) (Fig. 2).

In addition to the assays that were performed using pooled peptides, we evaluated epitope-specific IFNγ responses to individual peptides at 49 dpv, using PBMC from pigs in groups PigMatrix-EDV and FluSure (five from each group). Triplicate assays were performed for all peptide stimulations

and for controls. After 18 h of incubation in a 37°C humidified 5% CO_2 incubator, the ELISpot plates were washed and developed according to the manufacturer's recommendations. The ELISpot plates were then scanned in a CTL-ImmunoSpot S5 UV analyzer and spot counts were recorded using the ImmunoSpot software (Cellular Technology Ltd., Shaker Heights, OH). Results were recorded as the average number of spot forming cells (SFC) over background and adjusted to spots per 10^6 PBMCs. A response was considered positive if the number of spots was greater than or equal to 20 SFC over background per 10^6 PBMC.

At the end of the study, pigs were SLA-typed using a low-resolution group-specific typing method [31] to evaluate SLA diversity and correlate epitope predictions with IFNγ responses. Select pigs were typed: two pigs from group NV, eight from Sham, seven from PigMatrix-EDV, seven from FluSure.

Retrospective analysis

IFNγ responses to individual peptides were compared to predictions using cohort-specific class I and II SLA PigMatrices. Class I peptides were scored and considered potential binders if the mean of significant Z-scores was above 1.64. Class II peptides with cluster scores above 10 were categorized as potential ligands. Experimentally, peptides that induced more than 20 SFC over background per 10^6 PBMCs in at least one pig were considered positives. Based on the comparison of experimental results and predictions, peptides were divided into one of four categories (true positives, true negatives, false positives, and false negatives). True-positive peptides were predicted and validated *in vitro* as immunogenic, while true-negative peptides were predicted and biologically validated to be non-immunogenic. False negative peptides were predicted to be non-immunogenic, yet produced a positive response; false positive peptides were predicted to be immunogenic, but produced no response in the IFNγ ELISpot assay. To evaluate the predictive performance of the matrices, we calculated the positive and negative predictive values (PPV and NPV, respectively) and area under the receiver operating characteristic (ROC) curve (AUC) using the sensitivity and 1 - specificity (false positive rate) values.

Antibody evaluation

Pig serum was collected at 0 and 42 dpv for hemagglutination inhibition (HI) assay to assess antibody responses following vaccination as described previously [43]. Briefly, sera were heat-inactivated at

56°C for 30 min and then treated with a 20% suspension of kaolin (Sigma-Aldrich, St. Louis, MO) and subjected to adsorption with 0.5% turkey red blood cells (RBC) to remove nonspecific hemagglutinin inhibitors and natural serum agglutinins. The HI assays were then performed using H1N1pdm09 and OH/07 (γ-cluster H1) strains as antigen. Titers were determined using two-fold serial dilutions to detect the reciprocal endpoint of HI, $\log_2$ transformed and reported as the average geometric mean reciprocal titer for each group. Sera with titers <40 were considered HI negative or suspect.

Statistical analysis

IFNγ responses to restimulation treatments (pooled peptides and WV) in the PigMatrix-EDV group and the FluSure group, measured at 42, 49 and 63 dpv, were compared using a Kruskall-Wallis test followed by side by side comparisons of the groups using Dunn's correction for multiple comparisons. The same test was used for comparison of HI antibody titers between groups at 42 dpv. Wilcoxon matched-pairs test was used to compare IFNγ responses within groups. To evaluate IFNγ responses to more than two restimulation treatments for a group at a specific timepoint and the effect of the boosts in the PigMatrix-EDV group, the Friedman test using Dunn's correction was used. P values of less than 0.05 were considered significant. All the statistical analyses were performed using GraphPad Prism software (GraphPad, San Diego, CA).

Results

Epitope selection

A total of 28 class I and 20 class II peptides were down-selected for inclusion in the prototype PigMatrix-EDV IAV vaccine (Fig. 3), following immunoinformatic predictions. Peptides were selected based on predicted binding to class I (SLA-1*0101, 1*0401, 2*0101, and 2*0401) and class II (SLA-DRB1*0101, 0201, 0401, and 0601) SLA alleles.

Since external proteins (HA and NA) are highly variable, it was difficult to identify highly conserved potential epitopes. For this study, the minimum IAV strain coverage required for epitopes derived from HA and NA proteins was 25%. We selected epitopes to achieve the broadest possible coverage despite this constraint. In contrast, internal proteins are conserved due to the presence of the

TRIG cassette; therefore, the coverage threshold for peptides selected from internal proteins (M1, M2, NP, NS1, NS2, PA, PB1, PB1-F2, PB2) was 85%.

Fourteen class I peptides were selected from external proteins and 14 peptides were selected from internal proteins. The mean Z-score of class I peptides was 2.87(1.03), [reported as mean(standard deviation)]; these are high-scoring peptides that are considered likely to be T cell epitopes. Twenty-four of the class I peptides (85.7%) were predicted to bind to four alleles. Eleven of the 14 class I peptides (78.6%) identified in the external proteins were >85% identical in at least three of seven IAV strains. Similar epitopes were selected to evaluate strain specificity; HA_1 and HA_2 differed by one amino acid, but HA_1 was 100% identical in four strains, whereas HA_2 was present only in one IAV (OH/07). NA_14 was identified in one IAV, but its sequence was 77.8% identical (7 of 9 amino acids) to NA_11, which was conserved in two other IAVs. Both peptides were predicted to bind to four class I SLA alleles. For the putative class I peptides derived from internal proteins, 11 of 14 (78.6%) were 100% identical in the IAVs analyzed.

Cluster scores for all the selected class II ICS were greater than 10. Eight of the 20 ICS were derived from external and 12 from internal proteins. Their lengths ranged from 16 to 25 amino acids. All the peptides had at least one 9-mer frame predicted to bind to at least three SLA class II alleles; 80% (16 of 20 peptides) had at least one 9-mer predicted to bind to all four SLA class II alleles. From the external proteins, five of the predicted peptides were >85% identical in at least three IAV strains. Class II peptides derived from internal proteins were >85% identical in all seven IAVs, with exception of M_10 that had 84.2% identity (differed by 3 amino acids) with its counterpart in OH/04. Taken altogether, the immunoinformatics-predicted sequences represent a set of potentially broadly reactive swine influenza T cell epitopes.

Epitope-driven DNA vaccine construction

As a tool to evaluate epitope-specific responses to predicted peptides, we designed two prototype DNA vaccines; one containing class I-restricted epitopes and one containing class II-restricted epitopes (Supplemental Text 1). To minimize potential junctional immunogenicity of the class I construct, breakers (‘AAY’) were inserted at seven of 27 peptide junctions. In one case, a ‘spacers’ (‘GPGPG’) was introduced into the class II construct to disrupt the formation of junctional epitopes. Both

constructs were designed to avoid potential transmembrane domains. The DNA vaccine vectors also contained signal sequences to target the string of epitopes to the proteasome or the secretory pathway. These signal sequences, UbA76 for class I and TPA for class II, were of human origin; however, BLAST analysis showed that amino acid sequences from both were 99% and 71% identical, respectively, to their swine counterparts.

T cell immunogenicity

Epitope-specific responses to pooled PigMatrix-predicted peptides were demonstrated in immune recall IFNγ ELISpot assays using PBMC isolated at 42 (day of second boost), 49, and 63 dpv from animals in the PigMatrix-EDV and Sham groups (Fig. 4A). The four peptide pools (All, Int, Ext-II, and Int-I; Fig. 2) used for restimulation induced statistically significant different responses between pigs vaccinated with PigMatrix-EDV and Sham ($p<0.05$). IFNγ responses measured in pigs from PigMatrix-EDV and FluSure groups were significatively different ($p<0.05$), with exception of restimulation with pool Ext-II. No significant differences were observed between pigs vaccinated with FluSure and Sham. In PigMatrix-EDV-vaccinated pigs, we expected class II epitopes to dominate in the immune response to external proteins, and class I epitopes to dominate the immune response to the internal proteins. Contrary to our expectation, the number of IFNγ SFC induced by pools of class II peptides from external proteins (Ext-II) and class I peptides from internal proteins (Int-I) was below 20, which was significantly lower ($p<0.001$) than responses to all peptides pooled together (All) and peptides derived from internal proteins (Int). SFC were not statistically different between All and Int pools ($p=0.74$). These results suggest that immune responses to class II predicted epitopes contained in internal IAV proteins dominate the PigMatrix-EDV-induced response.

IFNγ SFC induced by restimulation with WV in PBMC from pigs in the PigMatrix-EDV group were also statistically different from Sham ($p<0.05$) at all three measured time points. This result suggests that T cells raised against epitopes contained in the prototype DNA vaccine recognize epitopes presented in whole virus stimulation *in vitro*.

It is interesting to note that the IFNγ SFC induced by restimulation with All and Int pools were not statistically different from responses to WV in PBMC from the PigMatrix-EDV group, with the exception of responses at 49 dpv (All: 49 (32.34), Int: 43.57 (33.93), WV: 84.29 (47.92) SFC per 10^6

PBMC; p=0.02; Fig. 4A). This may suggest that the epitopes in the peptide pools were recognized by T cells that were responsible for the majority of T cell responses to WV, *in vitro*. Differences in the antigen presentation processes (*in vitro*) for WV and peptides may explain the differences at 49 dpv [44]. Alternatively, WV RNA could have played a role in the expansion of T cell responses *in vitro* [45]. It is also possible that the IFNγ ELISpot assay only sampled a fraction of the antigen-specific cells present in the PBMC after *in vitro* stimulation; thus, technical limitations may explain comparable responses between pooled peptides and WV restimulation.

Interestingly, IFNγ responses to WV restimulation did not significantly differ in PigMatrix-EDV (73.84 (54.48)) and FluSure-immunized (127.55 (175.86)) pigs (Fig. 4A). Note that the mean and high variability in the FluSure group was due to consistently high recall responses in PBMC from one "high responder pig" (FS-442; Fig. 4A, marked with +) at the three time points tested (mean IFNγ SFC per 10^6 PBMC excluding this pig was 58.52 (51)). Thus, the epitope-based vaccine elicited consistent IFNγ responses equivalent to those induced by a tetravalent commercial WIV vaccine.

Boost immunizations in the PigMatrix-EDV group did not result in significant changes in the number of IFNγ SFC when PBMCs were restimulated with All, Int pools and WV (Fig. 4A and enhanced in Fig. 4B).

Restimulation with individual peptides

As noted before, we suspected that recall response to peptide pools in PBMC from PigMatrix-EDV vaccinated pigs was primarily driven by class II predicted epitopes derived from internal IAV proteins. This observation was confirmed by evaluating IFNγ responses to individual peptides a week after the second boost (49 dpv) (Fig. 4C and 4D). PBMC from pigs immunized with PigMatrix-EDV or FluSure (five from each group) were restimulated with individual class I and II peptides. Four class I peptides (derived from external proteins) and seven class II peptides (derived from internal proteins) induced more than 20 IFNγ SFC per 10^6 PBMC over background for at least one pig immunized with PigMatrix-EDV. At 49 dpv, none of the peptides induced significant responses in PBMC from the five FluSure-vaccinated pigs tested. Note that the high responder pig from the FluSure group, who registered the highest responses to peptide pool restimulation at 42, 49, and 63 dpv (FS-442, Fig. 4A, marked with +), was not included in the individual peptide restimulation assays.

Retrospective analysis using cohort-specific predictive matrices

Putative epitopes were predicted for binding to a set of SLA alleles prevalent in outbred swine populations [31,32]. To determine if those alleles were expressed in the study cohort, SLA types were determined at low resolution [31,32] at the end of the study for eight pigs from the Sham group, seven from PigMatrix-EDV, seven from FluSure, and two from NV group (Supplemental Table 2). By chance, none of the SLA-typed pigs tested in individual peptide ELISpot assays (Table 1), expressed any of the alleles used for epitope predictions. A pig that responded to four class I peptides (PigMatrix-EDV-427) was not SLA-typed; thus, no correlation of immune recall and epitope predictions could be made between the existing matrices and these ELISpot data.

To retrospectively evaluate the IFNγ responses to individual peptides and the association with specific SLA haplotypes, we developed class I and II matrices specific for the most frequent SLA-1, SLA-2, and SLA-DRB1 alleles expressed in the actual cohort (cohort-specific, Table 1). Although certain low-resolution results were ambiguous, we can make some assumptions based on common associations. For SLA class II, based on common association with DQB1 and DQA alleles [32], we expect that SLA-DRB1*0401-02 is likely to be DRB1*0402 and DRB1*06XX is likely to be DRB1*0602. These two alleles were expressed in 79% of the typed pigs. For the rest of the frequently expressed alleles, we developed XX01 as the default matrix (e.g. for DRB1*07XX, we developed SLA-DRB1*0701 prediction matrix). Thus, we developed cohort-specific prediction matrices for SLA-1*0801, 1*1201, 1*1301, 2*0501, 2*1201, DRB1*0402, 0602, 0701, and 1001.

The initial set of peptides was selected because they were predicted to bind promiscuously to the SLA alleles that are prevalent in outbred swine populations (SLA-1*0101, 1*0401, 2*0101, 2*0401, DRB1*0101, 0201, 0401, and 0601). However, a reduced number of peptides were predicted to bind to the actual, cohort-specific alleles, once this information was available (Fig. 5). For example, none of the peptides were predicted to bind the most frequently expressed SLA allele for this cohort (SLA-1*0801). Cohort-specific predictions yielded a total number of hits per allele, for this set of peptides, that was 41.7% lower than the initial prediction based on reported prevalent alleles. Despite the fact that the predictions did not correspond well with the sampled SLA, 23 PigMatrix-EDV peptides were still predicted to bind to alleles in the cohort, explaining the responses observed in the pool restimulation.

This also suggests that initial predicted promiscuity (i.e. ability for a peptide to bind to multiple alleles) present in selected peptides extends to additional cohort-specific alleles.

Cohort-specific class I predictions

Further evaluation of the cohort-specific predictions revealed that PigMatrix predictions for the 28 class I peptides (Fig. 5, top) had high sensitivity (1.0) and NPV (1.0), and moderate specificity (0.63). Cohort-specific prediction correctly classified 15 of the 24 peptides that were non-immunogenic. In terms of immunogenic peptides, four out of four were predicted as immunogenic, though nine false positive peptides were also observed, contributing to the low PPV (0.31).

To evaluate the predictive performance of the cohort-specific prediction, we built an ROC curve and then calculated the AUC (a value of 1 corresponds to a perfect prediction and 0.5 to a random prediction). The AUC was 0.81, which shows that predictions for cohort-specific alleles had high predictive power. However, the size of the dataset may have influenced these results; prospective studies on larger cohorts of pigs would be required to validate this observation.

Cohort-specific class II predictions

Predictions targeting cohort-specific SLA alleles showed that peptides had limited binding likelihood. Cohort-specific cluster scores were lower for 18 out of 20 peptides compared to cluster scores for SLA alleles reported as prevalent in outbred pigs (SLA-DRB1*0101, 0201, 0401, 0601; Fig. 5, bottom). Note that the cohort-specific cluster score of false negative peptide M_10, which calculated for four SLA-DRB1 alleles (DRB1*0402, 0602, 0701, and 1001), was below 10 (the threshold we set for potential binders), but it was still predicted to bind to the two SLA-DRB1 alleles (DRB1*0602 and 1001) expressed by PigMatrix-EDV SLA-typed pigs (pig PigMatrix-EDV-432 only expressed SLA-DRB1*1001), and corresponded to the positive responses observed.

Cohort-specific class II predictions (Fig. 5, bottom) had high sensitivity (0.86) and NPV (0.90), and moderate specificity (0.69) and PPV (0.60). They also showed high predictive power (AUC 0.77) for the set of 20 class II peptides. Nine of the 13 peptides that were non-immunogenic were correctly predicted and six out of seven immunogenic peptides were predicted as immunogenic. Only four peptides that were predicted to be immunogenic were non-immunogenic in T cell assays.

Overall, the retrospective analysis of 48 peptides using predictions for cohort-specific SLA alleles showed high sensitivity, moderate-to-high specificity and high predictive power for both class I and II SLA alleles. Predictions were particularly effective identifying non-immunogenic peptides as demonstrated by their high NPVs. Cohort-specific predictions correctly identified 24 out 37 non-immunogenic peptides and 10 out of 11 positive peptides. Still, it is important to mention that the limited number of peptides makes these results less robust. A larger dataset of peptides will be required to confirm the predictive power of the matrices.

Antibody responses

FluSureXP® induced detectable HI antibody against OH/07 γ-cluster virus, with no cross-reactivity to the H1N1pdm09 virus with serum collected 42 dpv (Table 2). HI antibody titers against OH/07 were significantly different from Sham and PigMatrix-EDV ($p<0.001$). PigMatrix-EDV did not induce a detectable positive HI antibody response against H1N1pdm09 or OH/07 at 42 dpv, which is not surprising, as the T cell epitopes were not expected to encode B cell epitopes.

Discussion

In this study, PigMatrix, an immunoinformatics tool for predicting swine T cell epitopes, was used to identify class I and II epitopes highly conserved among seven IAV strains representative of those prevalent in U.S. swine. To evaluate the immunogenic potential of the predicted peptides, IFNγ SFC recall responses were measured in pigs vaccinated intramuscularly with prototype DNA vaccines (PigMatrix-EDV) encoding strings of class I and II epitopes or a commercially available swine IAV vaccine. Recall responses induced by pooled peptides in PBMCs isolated from pigs vaccinated with PigMatrix-EDV were significantly greater than responses in pigs vaccinated with empty plasmids. Furthermore, PigMatrix-EDV-vaccinated pigs responded to WV (H1N1pdm09) restimulation, showing that the epitope-based immunization gave rise to T cells that are cross-reactive with epitopes present in the whole virus *in vitro*. In addition, overall responses to WV restimulation were comparable to those induced by All and Int pools. Moreover, epitope-specific recall responses to WV in pigs immunized with a prototype epitope-based vaccine were similar to responses in pigs immunized with the commercial vaccine.

A challenge study was conducted to evaluate protective efficacy of PigMatrix-EDV. Pigs were intranasally challenged with H1N1pdm09 virus, but due to age at challenge and route of challenge, pathology and viral load in non-vaccinates was limited, so assessing protection overall in vaccinates was also limited. There was no evidence of enhanced lesions (VAERD) in vaccinates, and outcome in DNA-vaccinates and FluSure vaccinates was similar (data not shown). Future work aimed at assessing efficacy of the DNA approach is warranted, and further consideration will need to be given to animal age, route of challenge, challenge strain, and SLA haplotype of animals to adequately evaluate the vaccine. Ideally, a group of pigs challenged with influenza A virus should be included to evaluate whether the predicted T cell epitopes are also induced during natural infection.

Our initial set of alleles used for T cell epitope prediction did not correspond well with the cohort ultimately selected. For this prospective study, we developed predictions for SLA alleles that had been reported to be frequently expressed in outbred swine populations [31,32]. However, post hoc SLA typing results showed that those alleles were not prevalent in pigs in the study. Still, some peptides induced IFNγ SFC responses, demonstrating the initial set of alleles positively predicted promiscuous epitopes. This has significant implications for vaccine design because identification of epitopes capable of binding to multiple SLA alleles limits the number of epitopes required to cover an SLA diverse population. A retrospective analysis using cohort-specific alleles showed that some of the peptides were predicted to bind to the new SLA alleles, although the set of peptides overall was not optimally matched to the cohort. These results indicate that selecting epitopes for promiscuity, when pig SLA-typing is not available, may be relevant because conservation of binding likelihood in a promiscuous epitope may extend to additional (untested) alleles. While it is clear that we will need to expand the set of alleles for future vaccine designs, this finding suggests that using immunoinformatics tools to identify promiscuous T cell epitopes can contribute to those future designs [26,46].

Designing epitope-based vaccines for pigs is hindered by the lack of information on SLA diversity in the U.S. swine population. A systematic evaluation of the SLA frequency will make it possible to develop and apply predictions for the most representative SLA alleles (supertypes) [47,48] to vaccine designs that cover a high percentage of the swine population. In addition, a more streamlined (i.e. rapid, high resolution, commercially available) approach to SLA typing would significantly

improve the ability to study T cell responses to influenza and other economically important diseases such as porcine reproductive and respiratory syndrome (PRRS) and porcine epidemic diarrhea (PED).

In this study, class II peptides from internal proteins were highly conserved (identity >85%) across all the analyzed strains and were shown to be the most immunogenic. Internal proteins from IAV are conserved across multiple strains because of the prevalence of two evolutionary lineages, H1pdm09 and TRIG, in the U.S. swine population [33]. We note that the genome sequences of the strains in the commercial vaccine are not available; however, it is likely that the internal epitopes were from the TRIG cassette (all seed strains predate introduction of H1N1pdm09 into the swine population). For this reason, it was interesting to see that PBMC from pigs immunized with FluSure had more limited IFNγ SFC responses to peptide pools, even though the pigs expressed similar SLA alleles to pigs in the PigMatrix-EDV group. This observation supports the hypothesis that epitope-based vaccines promote more efficient processing and presentation of their own epitopes as compared to whole-protein-based vaccines. Similar results were observed in mouse studies using T cell epitope-based DNA vaccines for *H. pylori*, where 33 out of 50 peptides stimulated more than 50 IFNγ SFC in splenocytes from the group vaccinated with an epitope-based DNA vaccine, but only two of the peptides were recognized in the group vaccinated with the whole bacteria lysate [49]. If epitope-based vaccines are able to induce immune responses to more individual epitopes than whole pathogen formulations containing the same epitopes, selection of the right epitopes, with the right breadth of SLA coverage, may lead to the development of more efficacious vaccines than currently exist [49].

Contrary to our expectations, we observed that IFNγ recall responses to class I peptides were restricted to *external* proteins, while responses to class II peptides were focused on epitopes derived from *internal* proteins. In human studies, most cross-reactive CD8+ (class I) and CD4+ (class II) T cell epitopes are derived from internal IAV proteins [14,50]. Compared to class I epitopes derived from internal proteins, HA- and NA-specific class I epitopes are said to be rare [51], but a few SLA-restricted HA and NA class I peptides have been reported [36,52]. In this study, the four class I peptides that induced IFNγ responses were derived from HA and NA. Sequence alignments using BLAST demonstrated that these peptides are conserved in different swine IAV strains. In humans, class II epitopes derived from HA and NA have also been reported [14,22,53], but none of the potential

epitopes from these antigens predicted by PigMatrix elicited measurable responses in the PigMatrix-EDV group (Fig. 4D). The seven class II peptides recognized by PBMC in this study were derived from internal proteins (M, NP, PA, and PB). Similar to class I peptides, these peptides are conserved in IAVs. Previous studies have shown that cross-reactive T cell responses to conserved epitopes may provide broader protection against diverse strains than antibodies that target variable antigens [15,22].

We searched the Immune Epitope Database (www.iedb.org) for swine influenza T cell epitopes and found that substrings of the predicted class II peptides NP_1, PA_7, PB_8, and M_12 have been reported to induce positive T cell responses, as measured by different methods (e.g. IFNγ ELISpot, tetramer staining, intracellular cytokine staining), for at least one human MHC class II allele. The published epitopes were derived from H1N1, H5N1, and H2N2 IAV strains. Thus epitopes that induce T cell responses in both human and pigs can be identified. Additionally, these epitopes may contribute to heterosubtypic cell-mediated responses against zoonotic IAV.

We did not expect that the epitope-based vaccine would induce antibodies, and indeed, PigMatrix-EDV did not induce HI antibodies that reacted to OH/07 or H1N1pdm09. While the commercial vaccine induced antibodies against OH/07, they did not cross-react with H1N1pdm09. The commercial vaccine contains four IAV strains (H1γ, H1δ1, H1δ2 H1N1 viruses, and one cluster IV H3N2 virus). OH/07 is an H1γ virus, which explains the positive HI response to this virus.

In conclusion, observed epitope-specific IFNγ recall responses demonstrate the potential for PigMatrix to predict conserved, promiscuous and immunogenic T cell epitopes. Further studies will evaluate the utility of PigMatrix for designing epitope-driven vaccines for swine. Epitope-driven T cell responses may not fully prevent IAV infection, but could reduce viral burden, as was observed for the 2009 H1N1 outbreak [22]. Rapid viral clearance and lower morbidity are important objectives for swine IAV vaccines, since current vaccines do not provide complete protection against variant strains. Moreover, epitope prediction tools could be used to assess the potential for existing commercial vaccine strains to protect against newly emergent strains of IAV. Improved immunoinformatics tools that target a comprehensive set of SLA alleles may contribute to the development of vaccines against other prominent swine diseases and provide a significant positive impact for pig health and swine producers.

Acknowledgments

The authors are grateful to Zahra Olson for her invaluable contribution in the performance of the experiments. We also thank Lilia Walther for her excellent technical assistance. Dr. Amy Vincent for providing accession numbers for MN/08 strain of IAV used in the analysis. Dr. Chris Bailey-Kellogg developed the concatemer optimization algorithm; we acknowledge his contribution and thank him for his help with the analysis. Thanks to Dr. Sam Ho for performing the SLA-typing. The authors also acknowledge the support of Jacob Tivin in the implementation of PigMatrix on iVAX. We also thank Ryan Tassone for his involvement in planning ELISpot assays.

References

1. Wideman G. Impact of Influenza A in Pork Production. Proceedings of the London Swine Conference. London, Ontario; 2013. p. 191–3.

2. Opriessnig T, Giménez-Lirola LG, Halbur PG. Polymicrobial respiratory disease in pigs. Anim Health Res Rev. 2011;12:133–48.

3. Kothalawala H, Toussaint MJM, Gruys E. An overview of swine influenza. Vet Q. 2006;28:46–53.

4. Vincent A, Awada L, Brown I, Chen H, Claes F, Dauphin G, et al. Review of Influenza A Virus in Swine Worldwide: A Call for Increased Surveillance and Research. Zoonoses Public Health. 2014;61:4–17.

5. Nelson MI, Vincent AL. Reverse zoonosis of influenza to swine: new perspectives on the human-animal interface. Trends Microbiol. 2015;23:142–53.

6. Anderson TK, Campbell BA, Nelson MI, Lewis NS, Janas-Martindale A, Killian ML, et al. Characterization of co-circulating swine influenza A viruses in North America and the identification of a novel H1 genetic clade with antigenic significance. Virus Res. 2015;201:24–31.

7. Vincent AL, Ma W, Lager KM, Janke BH, Richt JA. Swine influenza viruses a North American perspective. Adv. Virus Res. 2008;72:127–54.

8. Kitikoon P, Nelson MI, Killian ML, Anderson TK, Koster L, Culhane MR, et al. Genotype patterns of contemporary reassorted H3N2 virus in US swine. Journal of General Virology. 2013;94:1236–41.

9. Lorusso A, Vincent AL, Gramer ME, Lager KM, Ciacci-Zanella JR. Contemporary epidemiology of North American lineage triple reassortant influenza A viruses in pigs. Curr. Top. Microbiol. Immunol. 2013;370:113–32.

10. Chen Q, Madson D, Miller CL, Harris DLH. Vaccine development for protecting swine against influenza virus. Animal Health Research Reviews. 2012;13:181–195.

11. Braucher DR, Henningson JN, Loving CL, Vincent AL, Kim E, Steitz J, et al. Intranasal vaccination with replication-defective adenovirus type 5 encoding influenza virus hemagglutinin elicits protective immunity to homologous challenge and partial protection to heterologous challenge in pigs. Clin. Vaccine Immunol. 2012;19:1722–9.

12. Gauger PC, Vincent AL, Loving CL, Henningson JN, Lager KM, Janke BH, et al. Kinetics of Lung Lesion Development and Pro-Inflammatory Cytokine Response in Pigs With Vaccine-Associated Enhanced Respiratory Disease Induced by Challenge With Pandemic (2009) A/H1N1 Influenza Virus. Vet Pathol. 2012;49:900–12.

13. Loving CL, Brockmeier SL, Vincent AL, Gauger PC, Zanella EL, Lager KM, et al. Cross-fostering to prevent maternal cell transfer did not prevent vaccine-associated enhanced respiratory disease that occurred following heterologous influenza challenge of pigs vaccinated in the presence of maternal immunity. Viral Immunol. 2014;27:334–42.

14. Lee LY-H, Ha DLA, Simmons C, de Jong MD, Chau NVV, Schumacher R, et al. Memory T cells established by seasonal human influenza A infection cross-react with avian influenza A (H5N1) in healthy individuals. J. Clin. Invest. 2008;118:3478–90.

15. Price GE, Soboleski MR, Lo C-Y, Misplon JA, Pappas C, Houser KV, et al. Vaccination focusing immunity on conserved antigens protects mice and ferrets against virulent H1N1 and H5N1 influenza A viruses. Vaccine. 2009;27:6512–21.

16. Yap KL, Ada GL, McKenzie IF. Transfer of specific cytotoxic T lymphocytes protects mice inoculated with influenza virus. Nature. 1978;273:238–9.

17. Belz GT, Wodarz D, Diaz G, Nowak MA, Doherty PC. Compromised influenza virus-specific CD8(+)-T-cell memory in CD4(+)-T-cell-deficient mice. J. Virol. 2002;76:12388–93.

18. Kamperschroer C, Dibble JP, Meents DL, Schwartzberg PL, Swain SL. SAP is required for Th cell function and for immunity to influenza. J. Immunol. 2006;177:5317–27.

19. Marshall D, Sealy R, Sangster M, Coleclough C. TH cells primed during influenza virus infection provide help for qualitatively distinct antibody responses to subsequent immunization. J. Immunol. 1999;163:4673–82.

20. Lin YL, Askonas BA. Biological properties of an influenza A virus-specific killer T cell clone. Inhibition of virus replication in vivo and induction of delayed-type hypersensitivity reactions. J. Exp. Med. 1981;154:225–34.

21. McMichael AJ, Gotch FM, Noble GR, Beare PA. Cytotoxic T-cell immunity to influenza. N. Engl. J. Med. 1983;309:13–7.

22. Moise L, Tassone R, Latimer H, Terry F, Levitz L, Haran JP, et al. Immunization with cross-conserved H1N1 influenza CD4+ T-cell epitopes lowers viral burden in HLA DR3 transgenic mice. Hum Vaccin Immunother. 2013;9:2060–8.

23. Adar Y, Singer Y, Levi R, Tzehoval E, Perk S, Banet-Noach C, et al. A universal epitope-based influenza vaccine and its efficacy against H5N1. Vaccine. 2009;27:2099–107.

24. Goodman AG, Heinen PP, Guerra S, Vijayan A, Sorzano COS, Gomez CE, et al. A human multi-epitope recombinant vaccinia virus as a universal T cell vaccine candidate against influenza virus. PLoS ONE. 2011;6:e25938.

25. Moise L, Terry F, Ardito M, Tassone R, Latimer H, Boyle C, et al. Universal H1N1 influenza vaccine development: Identification of consensus class II hemagglutinin and neuraminidase epitopes derived from strains circulating between 1980 and 2011. Hum Vaccin Immunother. 2013;9:1598–607.

26. Gottlieb T, Ben-Yedidia T. Epitope-based approaches to a universal influenza vaccine. J. Autoimmun. 2014;54:15–20.

27. Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U, et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat. Biotechnol. 1999;17:555–61.

28. Terry FE, Moise L, Martin RF, Torres M, Pilotte N, Williams SA, et al. Time for T? Immunoinformatics addresses vaccine design for neglected tropical and emerging infectious diseases. Expert Rev Vaccines. 2014;1–15.

29. Gutiérrez AH, Martin WD, Bailey-Kellogg C, Terry F, Moise L, De Groot AS. Development and validation of an epitope prediction tool for swine (PigMatrix) based on the pocket profile method. BMC Bioinformatics. 2015;16:290.

30. De Groot AS, Jesdale B, Martin W, Saint Aubin C, Sbai H, Bosma A, et al. Mapping cross-clade HIV-1 vaccine epitopes using a bioinformatics approach. Vaccine. 2003;21:4486–504.

31. Ho C-S, Lunney JK, Franzo-Romain MH, Martens GW, Lee Y-J, Lee J-H, et al. Molecular characterization of swine leucocyte antigen class I genes in outbred pig populations. Anim. Genet. 2009;40:468–78.

32. Ho C-S, Lunney JK, Lee J-H, Franzo-Romain MH, Martens GW, Rowland RRR, et al. Molecular characterization of swine leucocyte antigen class II genes in outbred pig populations. Anim. Genet. 2010;41:428–32.

33. Anderson TK, Nelson MI, Kitikoon P, Swenson SL, Korslund JA, Vincent AL. Population dynamics of cocirculating swine influenza A viruses in the United States from 2009 to 2012. Influenza Other Respir Viruses. 2013;7 Suppl 4:42–51.

34. Lorusso A, Vincent AL, Harland ML, Alt D, Bayles DO, Swenson SL, et al. Genetic and antigenic characterization of H1 influenza viruses from United States swine from 2008. J. Gen. Virol. 2011;92:919–30.

35. Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Zaslavsky L, Tatusova T, et al. The influenza virus resource at the National Center for Biotechnology Information. J. Virol. 2008;82:596–601.

36. Zhang N, Qi J, Feng S, Gao F, Liu J, Pan X, et al. Crystal structure of swine major histocompatibility complex class I SLA-1 0401 and identification of 2009 pandemic swine-origin influenza A H1N1 virus cytotoxic T lymphocyte epitope peptides. J. Virol. 2011;85:11709–24.

37. De Groot AS, Terry F, Cousens L, Martin W. Beyond humanization and de-immunization: tolerization as a method for reducing the immunogenicity of biologics. Expert Rev Clin Pharmacol. 2013;6:651–62.

38. De Groot AS, Marcon L, Bishop EA, Rivera D, Kutzler M, Weiner DB, et al. HIV vaccine development by computer assisted design: the GAIA vaccine. Vaccine. 2005;23:2136–48.

39. Möller S, Croning MD, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics. 2001;17:646–53.

40. Velders MP, Weijzen S, Eiben GL, Elmishad AG, Kloetzel PM, Higgins T, et al. Defined flanking spacers and enhanced proteolysis is essential for eradication of established tumors by an epitope string DNA vaccine. J. Immunol. 2001;166:5366–73.

41. Livingston B, Crimi C, Newman M, Higashimoto Y, Appella E, Sidney J, et al. A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes. J. Immunol. 2002;168:5499–506.

42. Loving CL, Vincent AL, Pena L, Perez DR. Heightened adaptive immune responses following vaccination with a temperature-sensitive, live-attenuated influenza virus compared to adjuvanted, whole-inactivated virus in pigs. Vaccine. 2012;30:5830–8.

43. Gorres JP, Lager KM, Kong W-P, Royals M, Todd J-P, Vincent AL, et al. DNA vaccination elicits protective immune responses against pandemic and classic swine influenza viruses in pigs. Clin. Vaccine Immunol. 2011;18:1987–95.

44. Trombetta ES, Mellman I. Cell biology of antigen processing in vitro and in vivo. Annu. Rev. Immunol. 2005;23:975–1028.

45. Ulmer JB, Mason PW, Geall A, Mandl CW. RNA-based vaccines. Vaccine. 2012;30:4414–8.

46. De Groot AS, Bishop EA, Khan B, Lally M, Marcon L, Franco J, et al. Engineering immunogenic consensus T helper epitopes for a cross-clade HIV vaccine. Methods. 2004;34:476–87.

47. Southwood S, Sidney J, Kondo A, del Guercio MF, Appella E, Hoffman S, et al. Several common HLA-DR types share largely overlapping peptide binding repertoires. J. Immunol. 1998;160:3363–73.

48. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. BMC Immunology. 2008;9:1.

49. Moss SF, Moise L, Lee DS, Kim W, Zhang S, Lee J, et al. HelicoVax: epitope-based therapeutic Helicobacter pylori vaccination in a mouse model. Vaccine. 2011;29:2085–91.

50. Assarsson E, Bui H-H, Sidney J, Zhang Q, Glenn J, Oseroff C, et al. Immunomic analysis of the repertoire of T-cell specificities for influenza A virus in humans. J. Virol. 2008;82:12241–51.

51. Sun Y, Liu J, Yang M, Gao F, Zhou J, Kitamura Y, et al. Identification and structural definition of H5-specific CTL epitopes restricted by HLA-A*0201 derived from the H5N1 subtype of influenza A viruses. J. Gen. Virol. 2010;91:919–30.

52. Pedersen LE, Breum SO, Riber U, Larsen LE, Jungersen G. Identification of swine influenza virus epitopes and analysis of multiple specificities expressed by cytotoxic T cell subsets. Virol. J. 2014;11:163.

53. Babon JAB, Cruz J, Orphin L, Pazoles P, Co MDT, Ennis FA, et al. Genome-wide screening of human T-cell epitopes in influenza A virus reveals a broad spectrum of CD4(+) T-cell responses to internal proteins, hemagglutinins, and neuraminidases. Hum. Immunol. 2009;70:711–21.

Table 1. **Low resolution SLA-type alleles of pigs tested in individual peptide ELISpot assays.**

		SLA class I[a]		SLA class II[a]
Group	**Pig**	**SLA-1**	**SLA-2**	**DRB1**
PigMatrix-EDV	429	08XX,12XX,1301	0901-02,12XX	06XX,10XX
	431	08XX,12XX,1301	05XX,10XX	06XX,10XX
	432	08XX	05XX,10XX	10XX
	433	08XX,12XX,1301	0901-02,12XX	06XX,10XX
FS	435	1103,12XX,1301	10XX,jh02	06XX
	436	08XX	05XX,12XX	0401-02,10XX
	437	12XX,1301	10XX	06XX
	439	08XX	12XX	0401-02
	441	1103,12XX,1301	10XX,jh02	06XX

[a]Only loci for which prediction matrices were developed are shown.

Table 2. Geometric mean reciprocal titers of HI antibodies to different virus in sera collected at 42 dpv.

	Viral antigen*	
Vaccine group	H1pdm09	OH/07
NV	7	6
Sham	6	6
PigMatrix-EDV	10	6
FluSure	8	104

*Titers <40 are considered negative or suspect.

	Days post-initial vaccination (dpv)							
Vaccination group	**0**		**21**	**42**			**49**	**63**
Sham	Prime	HI assay	Boost 1	Boost 2	HI assay	IFNγ ELISpot	IFNγ ELISpot	IFNγ ELISpot
PigMatrix-EDV	Prime		Boost 1	Boost 2				
FluSure	Prime		Boost					
NV								

8 pigs/group

Fig. 1. Experimental outline.

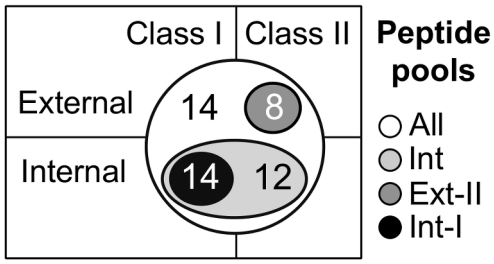


Fig. 2. Peptide pools tested. All: 48 peptides. Int: 26 class I and II peptides predicted from internal proteins. Ext-II: Eight class II peptides predicted from external proteins. Int-I: 14 class I peptides predicted from internal proteins.

CLASS I peptide	Peptide ID	H1N1			H1N2		H3N2	
		pdm09	IL/10	OH/07	MN/12	MN/08	TX/98	OH/04
GMVDGWYGY	HA_1							
GMIDGWYGY	HA_2							
SVKNGTYDY	HA_3							
NADTLCIGY	HA_4							
TSADQQSLY	HA_5							
ITIGKCPKY	HA_6							
RIYQILAIY	HA_7							
LSTASSWSY	HA_8							
GTIKDRSPY	NA_10							
EMNAPNYHY	NA_11							
ELDAPNYHY	NA_14							
KSCINRCFY	NA_12							
EICPKLAEY	NA_13							
DTVHDRTPY	NA_9							
LASCMGLIY	M_15							
DLLENLQAY	M_16							
SLLTEVETY	M_17							
GAKEVALSY	M_18							
NMDKAVKLY	M_19							
LTEVETYVL	M_20							
NTDLEALME	M_21							
VSDGGPNLY	NP_24		-					
GTEKLTITY	NP_25		-					
CTELKLSDY	NP_28		-					
AFDERRNKY	PA_26		-					
QVSRPMFLY	PB1_22		-					
DTVNRTHQY	PB1_27		-					
ASQGTKRSY	PB2_23		-					

CLASS II peptide	Peptide ID	H1N1			H1N2		H3N2	
		pdm09	IL/10	OH/07	MN/12	MN/08	TX/98	OH/04
GDKITFEATGNLVVPRY	HA_17							
YEELREQLSSVSSFER	HA_18							
VPRYAFAMERNAGSGIIIS	HA_19							
STRIYQILAIYSTVASSLVLV	HA_20							
CRTFFLTQGALLNDKH	NA_13							
NQTYVNISNTNFAAGQSVVSVKL	NA_14							
SVVSVKLAGNSSLCPV	NA_15							
MANLILQIGNIISIWISHS	NA_16							
TRQMVHAMRTIGTHPSSSA	M_10							
SCMGLIYNRMGTVTTEAAFGLVC	M_11							
TYVLSIIPSGPLKAEIAQRLESV	M_12							
IEDLIFLARSALILRGSVAHKSCLP	NP_1		-					
IDPFKLLQNSQVVSLMRP	NP_2		-					
TRGVQIASNENVETMDSNTLELR	NP_3		-					
FQDILMRMSKMQLGSSSE	NS_4		-					
FEQITFMQALQLLLEVE	NS_5		-					
EVHIYYLEKANKIKSEKTHIHIF	PA_6		-					
RSKFLLMDALKLSIEDP	PA_7		-					
MMGMFNMLSTVLGVSI	PB_8		-					
YRYGFVANFSMELPSFGVSG	PB_9		-					

100%	99-85%	<85%

Fig. 3. Class I and II predicted peptides. Peptides were selected based on predicted binding to class I and class II SLA alleles and conservation in IAV strains. The identity percentage between peptides and IAV strains is shown (100% dark gray, 99% - 85% gray, and <85% white). The Peptide ID is coded to the source protein. Sequences not available are marked with -.

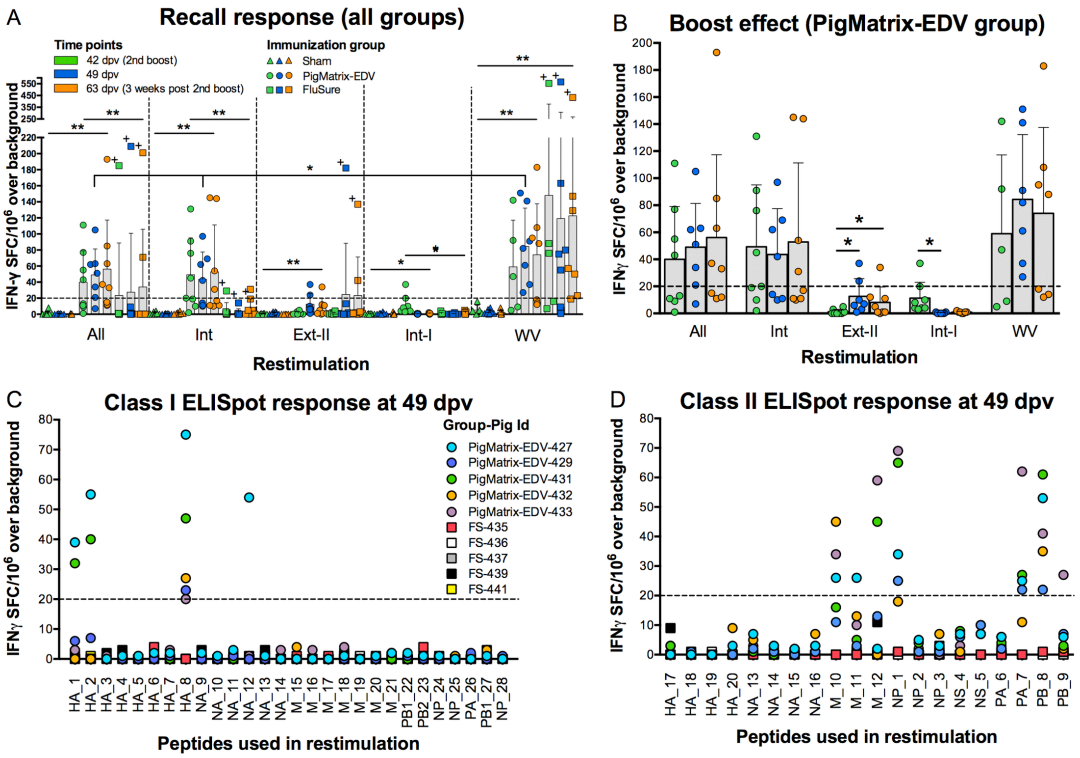


Fig. 4. Peptide immunogenicity measured by IFNγ ELISpot. (A) PBMC (2.5 x 10^5) isolated at three different time points (42, 49, and 63 dpv) from pigs immunized with empty plasmid (Sham), epitope-driven DNA vaccine (PigMatrix-EDV) and commercial vaccine (FluSure) were restimulated with pooled peptides (All, Int, Ext-II, and Int-I) at 10 µg/mL and whole virus (WV). The number of epitope-specific IFNγ spot forming cells (SFC) induced by the pools were measured using ELISpot assays. "High responder pig" (FS-442) is marked with +. (B) To evaluate vaccine boost effect, IFNγ responses to pooled peptides were measured at three different time points. For A and B, SFC over background, adjusted to spots per 10^6 of PBMC seeded, are represented with bars indicating means and error bars indicating standard deviation (SD). Pooled peptide responses showing statistical significance when compared to Sham are indicated: **p<0.01, *p<0.05. Significant statistical difference for PigMatrix-EDV between restimulations at 49 dpv is also shown. Same colors and shapes are used in both figures. (C) PBMC from pigs vaccinated with PigMatrix-EDV and FluSure were restimulated with individual class I peptides and (D) class II peptides one week after the second boost (49 dpv). For C and D, SFC over background per 10^6 PBMC are shown. A response was considered positive if the number of spots was greater than or equal to 20 SFC over background per 10^6 PBMCs (dashed line).

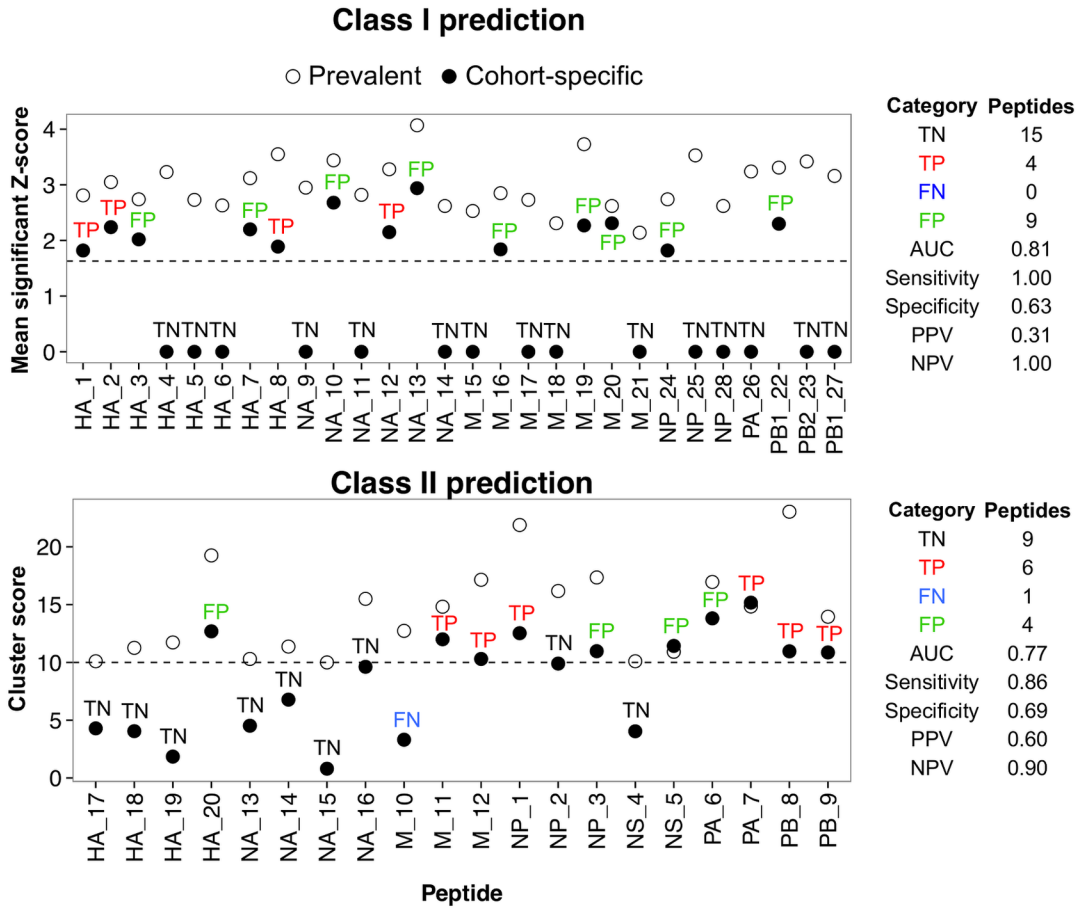


Fig. 5. Comparison between prediction for prevalent and cohort-specific SLA alleles. Peptides were predicted to bind to a set of previously reported class I and class II SLA alleles prevalent in the U.S. swine population (prevalent). Based on low-resolution SLA-typing results, those alleles were not represented in the studied pigs. Prediction matrices were developed to predict binding potential of peptides to the most frequent SLA alleles found in the cohort (cohort-specific). (Top) Mean of significant Z-scores (above 1.64) over prevalent class I SLA alleles (SLA-1*0101, 1*0401, 2*0101, and 2*0401) and cohort-specific (SLA-1*0801, 1*1201, 1*1301, 2*0501, and 2*1201) are shown for each peptide. Peptides with a mean of significant Z-scores above 1.64 (dashed line) are considered potential binders. (Bottom) Cluster scores calculated for prevalent class II SLA alleles (DRB1*0101, 0201, 0401, and 0601) and cohort-specific alleles (DRB1*0402, 0602, 0701, and 1001) are shown for each peptide. Cluster scores above 10 (dashed line) are considered as potential binders. Based on the retrospective evaluation, peptides were classified in four categories (TN: true negatives, TP: true positives, FN: false negatives, and FP: false positives). AUC, Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) are shown.

CHAPTER 4

Manuscript prepared for submission to Influenza and Other Respiratory Viruses

T cell epitope Content Comparison (EpiCC) of swine H1 influenza A virus hemagglutinin

Andres H. Gutiérrez[1], Vicki Rapp-Gabrielson[2], Crystal Loving[3], Frances E. Terry[4], Leonard Moise[1,4], William D. Martin[4], Anne S. De Groot[1,4]

[1]Institute for Immunology and Informatics, Department of Cell and Molecular Biology, University of Rhode Island, Providence, RI, USA

[2]Zoetis Inc., Kalamazoo, MI, USA

[3]Virus and Prion Diseases Research Unit, NADC, USDA ARS, Ames, IA, USA

[3]EpiVax Inc., Providence, RI, USA

*Corresponding Author: Anne S. De Groot, M.D.
Institute for Immunology and Informatics
Department of Cell and Molecular Biology
University of Rhode Island
Providence, RI, 02903, USA
E-mail: dr.annie.degroot@gmail.com

Abstract

Background: Hemagglutinin (HA) is the most variable antigen of swine influenza A viruses (IAV) and the major target of protective antibody responses; hemagglutination inhibition (HI) antibody titers are widely used to predict vaccine cross-protection. T cell epitopes that are cross-conserved between IAV strains also play a role in protection.

Objective: Determine the association of T cell immunogenic potential with vaccine-induced protection.

Methods: We developed a method for T cell epitope content comparison (EpiCC) and used it to compare field H1 swine IAV HAs to HAs in a commercial vaccine. The PigMatrix algorithm was used to predict T cell epitope content for twenty-three HA sequences representing the major H1 swine IAV phylo-clusters. The T cell epitope-based relatedness (EpiCC scores) between vaccine and field viruses was calculated and compared to experimental data from previous efficacy studies where immunization induced protection in the absence of cross-reactive HI antibodies.

Results: A comparison of HA T cell epitope content of the γ-cluster H1 vaccine virus to viruses used in homosubtypic challenge studies yielded a relatedness score associated with protection, assuming minimal epitope variation of other viral antigens.

Conclusion: The T cell epitope content relationship deduced from vaccine efficacy studies evaluated here may support the hypothesis that T cells contribute to vaccine efficacy. EpiCC could be used retrospectively and prospectively to estimate the immunologic relationship of genetically and antigenically variable viruses assessed in vaccine efficacy studies. EpiCC may complement studies of HI cross-reactivity and phylogenetic data for the prospective selection of influenza strains in flu vaccine development. Complete viral proteome EpiCC analysis would provide more comprehensive information to estimate vaccine efficacy.

Keywords: Computational immunology, hemagglutinin, SLA, swine influenza H1 viruses, vaccine efficacy, T cell epitope content comparison, T cell epitope prediction.

Introduction

Influenza A viruses (IAVs) cause a highly contagious respiratory disease in swine that has a significant economic impact on the pork industry [1]. Hemagglutinin (HA) and neuraminidase (NA) are the two major surface glycoproteins that define IAV subtypes and play a key role in antigenicity, pathogenesis, host range, and diagnosis. The segmented IAV genome allows for antigenic shift by reassortment of RNA segments from different viral strains infecting the same cell, generating novel viruses [2]. Antigenic drift due to accumulation of mutations in HA and NA also contributes to the remarkable diversity of IAVs [2].

Currently, H1N1, H1N2 and H3N2 are the predominant IAV subtypes co-circulating in the North American swine population. These subtypes are further subdivided based on the genetic and antigenic properties of HA. For H1 viruses, seven distinct genetic phylo-clusters (α, β, γ, γ-2, δ1, δ2 and pandemic (pdm09)), have been identified [3]. The HA gene of α, β γ, and pdm09 cluster viruses is most similar to classical swine H1N1 (cH1N1) [3]. HA from human-origin δ viruses can be differentiated in two subclusters, δ1 and δ2 [4]. Based on the antigenic properties of HA, serological cross-reactivity between H1 clusters similar to cH1N1, can be variable, but there is limited to no cross-reactivity between these viruses and the even more divergent δ cluster viruses [3,5]. This diversity presents a challenge to pork producers, veterinarians and vaccine manufacturers who wish to match vaccine strains to relevant field strains.

Most of the North American commercial vaccines against swine IAV contain inactivated viruses. The predominant antibody responses induced by these vaccines is to the HA proteins [1,6]. Therefore, hemagglutination-inhibiting (HI) antibody titers as well as genetic information of HAs are used to evaluate the potential for one vaccine to protect against variant strains [6]. However, based on lung lesions and viral titers, full or partial protection of pigs can be observed following vaccination with inactivated commercial vaccines, in the absence of antibody cross-reactivity [1,6–12].

Protection in the absence of neutralizing antibodies is presumed to be due to cell-mediated immunity (CMI) to cross-conserved T cell epitopes [13,14]. Cross-conserved T cell epitopes have been associated with protection against influenza in other contexts. For example, in humans, immunity to cross-conserved epitopes during the 2009 H1N1 IAV pandemic may have contributed to attenuation of

morbidity in some age groups [13,15]. In a previous study, we showed that a DNA vaccine based on eight HA T cell epitopes and one NA epitope conserved between seasonal and pdm09 lowered lung viral loads in HLA-DR3 transgenic mice challenged with pdm09 [16]. Cross-protection induced by conserved antigens does not induce complete protection against infection, but reduces mortality, morbidity, virus replication, and viral shedding [17,18]. CMI depends on the activation of cytotoxic T lymphocytes (CTL, CD8) and T helper (Th, CD4) lymphocytes when their T cell receptors (TCR) recognize T cell epitopes presented by class I or class II major histocompatibility complex (MHC) molecules on the surface of antigen presenting cells or infected cells [19]. CTL kill virus-infected cells, whereas Th responses support memory CD8 T cell development and provide help to B cells for high-affinity, neutralizing antibody responses [13].

Here, we describe an approach for T cell **epi**tope **c**ontent **c**omparison (EpiCC) that compares putative class I and II epitopes to assess relatedness across antigens and predict potential vaccine efficacy. Using this method, we evaluated whether T cell epitope relatedness could explain protection against heterologous IAV challenge (in the absence of cross-reactive antibodies). EpiCC uses PigMatrix, an algorithm that predicts class I and II T cell epitopes specific to swine MHC (Swine Leukocyte Antigen, SLA) alleles [20]. PigMatrix and additional tools in the iVAX toolkit [21] have been used to prospectively identify conserved T cell epitopes from different H1 phylo-clusters [21,22]. We applied PigMatrix and EpiCC to HA proteins from 23 swine IAV strains representing the major H1 phylo-clusters circulating in the North American swine population. Since the internal genes in North American swine influenza have remained prevalent and conserved since 1998 (due to the emergence of the triple reassortant internal gene (TRIG) cassette), we assumed minimal T cell epitope differences therein and focused on the critical and most variable swine IAV antigen, HA [3,23,24]. Comparing the results of vaccine efficacy studies with the IAV EpiCC scores, we identified a level of T cell epitope related for a γ-cluster H1 vaccine virus associated with protection.

Methods

Sequences

HA sequences from 23 H1 IAV strains were included in the analysis (Table 1). Twenty sequences were from swine H1 viruses representing α, β, γ, γ-2, δ1, δ2 and pdm09 phylo-clusters and three were from the H1 vaccine viruses (γ, δ1 and δ2 H1 phylo-clusters) in FluSure XP® (FS; Zoetis Inc, Florham Park, NJ). Phylogenetic analysis was performed using MEGA 7 [25]. HA nucleotide sequences were aligned with MUSCLE and an evolutionary tree was inferred using the Maximum likelihood method with 500 bootstrap replicates.

MHC binding prediction

The HA amino acid sequences of the 23 IAV strains were screened using PigMatrix [20]. PigMatrix parses sequences into 9-mers and assesses the binding potential of each 9-mer *i* to SLA class I and II alleles. For each individual allele ***a*** in a set of MHC alleles ***A***, PigMatrix raw scores *r* are normalized to Z-scores using the average μ and the standard deviation σ of scores calculated for 100 000 random 9-mers.

$$Z(i)_a = \frac{(r - \mu)}{\sigma}$$

In this normalized set of scores for each SLA allele, 9-mers with Z-scores above 1.64 comprise the top 5% of sequences with significant SLA binding potential. Increasing Z-scores correlate with higher MHC binding probability.

The distribution of SLA alleles among pig herds in the United States is unknown. Binding was therefore predicted to a set of SLA class I and II alleles that were frequently expressed in a cohort tested in a previous study (SLA-I: SLA-1*0801, 1*1201, 1*1301, 2*0501, 2*1201, 3*0501, 3*0601, and 3*0701; SLA-II: SLA-DRB1*0201, 0402, 0602, 0701, and 1001) [22].

Since data are lacking on breadth of coverage for SLA, a comparison was performed using HLA alleles. The breadth of coverage of HLA alleles (i.e. the ability to cover a global population) is known [26,27]; therefore, the HLA-restricted T cell epitopes that could be identified in these IAV sequences was quantified and compared to the epitopes uncovered in the sets of SLA alleles. The following HLA class I and class II supertype alleles were used for this comparison, HLA-I: A*0101,

A*0201, A*0301, A*2402, B*0702, and B*4403, and HLA-II: DRB1*0101, 0301, 0401, 0701, 1101, 1301, and 1501 [28].

T cell epitope content comparison

The EpiCC algorithm assesses the relatedness of a protein sequence of a strain *s* and a protein sequence of a vaccine strain *v* based on a comparison of their T cell epitope content, predicted for the same set of MHC alleles *A*. For any comparison, the predicted epitope content can be either shared between sequences or unique to the strain or to the vaccine. Thus, the score of the comparison between the epitope content of *s* and *v* (EpiCC score or T cell epitope-based relatedness) considers scores of shared and unique epitopes (Fig. 1).

Intuitively, the epitope content of a protein depends on its epitope density. So, if a 'high epitope density' protein is compared to a highly similar protein and many of their epitopes are conserved or shared between the two strains, the scores of shared epitopes will be high; consequently, the score of the comparison of their epitope content (EpiCC score) will also be high. The score will be even higher if the 9-mer epitopes are predicted to bind with high probabilities to alleles in the set *A*.

We hypothesized that if epitopes in a vaccine match the epitopes in the challenge strain, and vaccine and strain unique epitopes are rare, the memory T cells induced by the vaccine are likely to cross-react with epitopes from the challenge strain. The model assumes (1) a naive immune system, (2) the vaccine does not induce memory T cells to challenge strain unique epitopes, and (3) the immune response to the vaccine could be biased to vaccine unique epitopes [29]. Consequently, the EpiCC score of two sequences is improved by the presence of shared epitopes and is lowered when strain and vaccine unique epitopes are present.

The EpiCC score is calculated using T cell epitope predictions for *s* and *v*. Each 9-mer $i \in s$ is compared to a corresponding 9-mer $j \in v$. The pairs of 9-mers *i,j* are determined from a local alignment of *s* and *v* sequences using the Smith-Waterman algorithm from EMBOSS [30]. For *i,j* where one of the 9-mers has a gap in position one, that 9-mer is considered "nonexistent", i.e. excluded from comparison.

For each *i,j* and each allele $a \in A$, the score of a shared T cell epitope *S(i,j)a* is computed only for cross-conserved epitopes (i.e. *i,j* with identical residues that face the TCR and predicted to bind to

allele *a*). We reasoned that epitopes with identical TCR-facing residues (TCRf), that are also predicted to bind to the same MHC allele, are more likely to induce cross-reactive memory T cells. This is a simple assumption because a TCR can recognize peptides with different TCRf [31], but it is a conservative initial approach to define potential cross-reactive epitopes [32]. For class I T cell epitope comparison, we assumed that *i* and *j* are cross-conserved, and potentially cross-reactive, if they have identical residues in positions 4, 5, 6, 7, and 8 and are predicted to bind to *a*, regardless of differences on their MHC-facing amino acids. For class II, amino acids in positions 2, 3, 5, 7, and 8 were considered TCRf. Positions were selected based on published analysis of peptide-MHC-TCR crystal structures [33]. *S(i,j)a* is calculated using predicted binding probabilities as follows:

$$S(i,j)_a = p(i)_a \cdot p(j)_a$$

where *p* is the cumulative probability in the Normal distribution for the Z-score. Since the binding of *i* and *j* to allele *a* are independent, *S(i,j)a*, the probability of them both occurring, is the product of the probabilities of each occurring (i.e. joint probability).

The score of a unique epitope *U(i,j)a* is determined for non-cross-conserved epitopes based on binding probabilities according to these criteria:

- Score of a strain unique epitope: $if\ Z(i)_a > 1.64 \rightarrow U(i,j)_a = p(i)_a$
- Score of a vaccine unique epitope: $if\ Z(j)_a > 1.64 \rightarrow U(i,j)_a = p(j)_a$

Note that for any given *i,j*, predicted epitopes cannot be both shared and unique for allele *a*, but they can be both strain unique and vaccine unique if *i* and *j* are predicted to bind allele *a*, and their TCRf are distinct. For *i,j* where both 9-mers are not predicted to bind allele *a*, *S(i,j)a* and *U(i,j)a* are undetermined.

Since the alleles in *A* are distinct, they are treated independently; hence, the score of shared epitopes for *i,j* over the full set of alleles can be calculated as a joint probability (i.e. product of the shared binding probabilities for individual alleles). However, given that the score of a shared epitope is calculated only for *i,j* where both 9-mers are predicted to be binders, the joint probability over multiple alleles underweights shared promiscuous epitopes. For this reason, we computed the sum of the probabilities instead.

For the calculation of the EpiCC score, we assume that binding of each 9-mer epitope is mutually exclusive and uniform. Thus, *E* is the sum of shared and unique epitope scores of each *i,j* normalized by the total number of compared pairs *p* to account for variable epitope densities, and by the number of MHC alleles in *A* allowing for comparison of values of *E* determined using different numbers of MHC alleles. Formally, the EpiCC score for sequences from a vaccine and strain is computed as:

$$E(s,v)_A = \frac{1}{|p| \cdot |A|} \sum_{i \in s; j \in v} \sum_{a \in A} S(i,j)_a - U(i,j)_a$$

The sum of class I and II $E(s,v)_A$ is the total epitope-based relatedness score for *s* and *v*. Note that *U(i,j)a* functions as a penalty; therefore, if $U(i,j)_a > S(i,j)_a$, $E(s,v)_A$ is negative.

Comparison of the predicted epitope content of any sequence to itself is considered its baseline EpiCC score ($E(s,s)_A$; Fig. 1) and it represents the predicted epitope density of the sequence and the binding probabilities of its epitopes. It follows that the maximum value of $E(s,v)_A$ can only be less than or equal to $E(v,v)_A$ or $E(s,s)_A$. For vaccines with low epitope content, $E(v,v)_A$ will be low, and $E(s,v)_A$ will be also low, even if *s* and *v* epitopes are highly similar. Thus, low $E(s,v)_A$ can be due to low epitope content of one or both sequences and/or low epitope relatedness between strains.

HA baseline EpiCC score comparison

We calculated the baseline EpiCC score of the HA sequence of each viral strain ($E(s,s)_A$). So as to evaluate whether the selection of MHC alleles had an effect on the baseline EpiCC scores, $E(s,s)_A$ was calculated using the epitope content predicted with four different sets of MHC alleles *A* (SLA-I, SLA-II, HLA-I, and HLA-II).

Comparison of HA T cell epitope content between field and vaccine viruses

We compared the epitope content (predicted using SLA alleles) of each HA to that of FS vaccine viruses ($E(s,v)_A$). Shared and unique class I, class II and total EpiCC scores were determined. We also explored the relationships between protein identity and EpiCC scores using regressions. To represent the presumed lower end of the identity spectrum, we analyzed HA from A/swine/North Carolina/A01442548/2012 (H3N2) and A/swine/Missouri/A01727926/2015 (H4N6) viruses (GenBank accession KC445235 and KU641621, respectively); their HA amino acid sequences had identities

between 41.1% and 43.6% when compared to HA from FS viruses. A random sequence that had the same number of amino acids as the average HA sequence in this data set and the average amino acid frequencies of the proteins in the Swiss-Prot database, was also included in the comparison.

Relationship between EpiCC scores and vaccine efficacy

Based on reports of the experimental outcomes of previously published challenge studies and challenge studies performed by Zoetis evaluating the efficacy of the FS H1γ vaccine virus, we analyzed whether a certain level of T cell epitope relatedness of HA was associated with protection and could explain protection observed without cross-reactive antibodies, assuming minimal variation of T cell epitope content among internal proteins. FS was considered protective if it reduced the percentage of lung lesions and viral titers in nasal swabs (i.e. nasal shedding) and/or in lung or lung lavage. We defined a total EpiCC score threshold to predict protection as the lowest EpiCC score for the comparison between the FS H1γ vaccine virus and challenge viruses, where studies demonstrated that the vaccine was protective.

Statistical analysis

Wilcoxon matched-pairs signed rank test was used to compare baseline EpiCC score of HA sequences defined using different sets of MHC alleles (e.g. $E(s,s)_{SLA\text{-}I}$ vs. $E(s,s)_{SLA\text{-}II}$; $E(s,s)_{SLA\text{-}I}$ vs. $E(s,s)_{HLA\text{-}I}$). The same test was applied to evaluate differences between SLA class I and II EpiCC scores for each vaccine virus. Correlation between class I and II baseline EpiCC scores was determined using the nonparametric Spearman's correlation coefficient (ρ). Correlation of class I and II EpiCC scores was evaluated using the same test. The correlation between identity and EpiCC scores was determined by Pearson correlation (r). P values (p) less than 0.05 were deemed significant. Analyses were performed using R 3.3.1 (Foundation for Statistical Computing, Vienna, Austria).

Results

HA baseline EpiCC scores, $\boldsymbol{E(s,s)_A}$

HA sequences from a range of swine IAV phylo-clusters (Fig. 2) were analyzed. Although vaccines for influenza contain many antigens, the critical antigen (and the most variable) is HA. Thus, for this analysis, we assumed HA as the main source of T cell epitope variability among viral strains. Across

the 23 H1 viruses, HA SLA class I baseline EpiCC scores, $E(s,s)_{SLA\text{-}I}$, were significantly lower ($p<0.001$) and less variable (0.049 (0.001); mean (standard deviation)) than class II, $E(s,s)_{SLA\text{-}II}$, (0.068 (0.004)) (Supplemental Fig. 1, SLA alleles), and they were not significantly correlated ($\rho= 0.18$, $p=0.19$). HA proteins from recently reported H1δ1 cluster viruses had the highest class II and total baseline EpiCC scores.

To evaluate the effect of allele specificity and the breadth of coverage of the set of SLA alleles, we compared the baseline EpiCC scores predicted using SLA alleles to that predicted using supertype HLA alleles. Class I baseline EpiCC scores for HLA alleles, $E(s,s)_{HLA\text{-}I}$, (0.063 (0.002)) were significantly higher ($p<0.001$) than those predicted using SLA alleles, $E(s,s)_{SLA\text{-}I}$ (0.049 (0.001); Supplemental Fig. 1). $E(s,s)_{HLA\text{-}II}$ (0.065 (0.004)) were significantly lower ($p<0.05$) than $E(s,s)_{SLA\text{-}II}$ (0.068 (0.004)). The lower baseline scores observed for SLA class I alleles as compared to HLA class I alleles may indicate that the set of SLA-I alleles selected for this study was not as broad in terms of population coverage and might not capture all the T cell epitope differences between strains. Alternatively, the strains may contain fewer epitopes that bind to SLA class I alleles, which is similar to the significantly lower ($p=0.04$) HLA class I baseline EpiCC scores compared to those of HLA class II using supertype HLA alleles. This result suggests that it is possible that for these sets of HAs and MHC alleles, class I epitope content is lower than that of class II.

Comparison of HA T cell epitope content between field and vaccine viruses

We then compared the SLA class I and II epitope content predicted for HA of each field virus to that of the vaccine viruses. Intuitively, HA proteins from similar strains will have similar epitope content. We observed that HA sequences from viruses within the same H1 cluster or in a cluster of the same HA lineage (cH1N1 or human seasonal) had higher scores for class I and II shared epitopes and lower scores for unique epitopes than viruses in clusters from a different HA lineage (Fig. 3). It is noteworthy that there were shared epitopes in all comparisons, even when comparing viruses from different HA lineages.

Scores of shared, strain unique and vaccine unique SLA class II epitopes were significantly higher than those for class I ($p<0.001$ for the three vaccine viruses), with the exception of scores of H1γ FS vaccine unique epitopes ($p=0.05$). Likewise, using HLA supertype alleles, scores of class II shared

epitopes were also significantly higher than those for class I ($p<0.01$) (Supplemental Fig. 2). Class II scores of unique epitopes were also higher for OK08 H1δ1 FS (vaccine unique) and NC05 H1δ2 FS (strain unique). Although the population coverage of the set of SLA class I may be limited, the scores calculated using HLA supertype alleles supports that it is possible that the analyzed HA sequences had lower class I than class II scores. Regardless, the comparison between strains was performed using identical sets of SLA class I and II alleles, normalizing the comparisons.

Using radar plots, it is possible to visualize differences between the EpiCC scores of each vaccine virus compared to each field virus. In Fig. 4, each axis corresponds to one virus HA sequence. The HA sequences are sorted clockwise by nucleotide identity, relative to the IA00 H1γ FS virus. The radar plots show that nucleotide identity did not exactly correlate with EpiCC scores (e.g. IA12 H1γ-2 and IA15 H1γ). The highest scores on each of the plots is $E(v,v)_A$ (i.e. vaccine compared to itself). HA sequences of viruses with the same H1 cluster of the vaccine registered the highest EpiCC scores. Note that for comparisons where the score for unique epitopes was greater than the score for shared epitopes, EpiCC scores were below zero. For each vaccine virus class I and II EpiCC scores were significantly different from each other ($p<0.05$).

Class I and II EpiCC scores correlated with the HA amino acid sequence identity of vaccines and viral strains ($r=0.86 – 0.89$). However, the relationship between identity and EpiCC scores was nonlinear second order polynomial (Fig. 5; $R^2=0.94 – 0.98$). EpiCC scores for sequences that had approximately 40% identity were no different from EpiCC scores for random amino acid sequence of similar length. Similar results were observed for correlation with HA nucleotide sequence identity (Supplemental Fig. 3). Unlike identity, EpiCC scores account only for differences in amino acids involved in predicted T cell epitopes. For example, there are only 35 amino acids that differ between the vaccine strain IA00 H1γ FS and the field strain CA09 H1pdm, but only 16 of these amino acids were involved in putative SLA binders and had an effect in the class II EpiCC score. Furthermore, only four of the residues were contained in 9-mers that were predicted to bind to three or more class II SLA alleles.

Relationship between EpiCC scores and vaccine efficacy

We compared EpiCC scores with results of reported vaccine efficacy studies for the FS γ-cluster vaccine strain (IA00 H1γ FS) against heterologous viruses representing α, β, γ, or H1pdm clusters (Table 2) to evaluate whether certain level of HA T cell epitope relatedness between vaccine and challenge strains was associated with protection. The primary measure for assessing vaccine efficacy in these studies was reduction of lung lesions, with reduction of viral nasal shedding and/or virus titers in lung or lavage fluid at necropsy as secondary outcomes. A vaccine was considered protective in our analysis if there was a reduction of macroscopic pneumonia and virus titers in nasal swabs and/or in lung specimens collected at necropsy. If the vaccine significantly reduced virus titers, but not lung lesions, it was considered partially protective. Using these data, we found a total EpiCC score (sum of class I and II EpiCC scores) threshold associated with protective efficacy for this set of challenge studies.

For the six vaccine efficacy studies considered in this analysis, FS conferred protection against challenge with five different H1 cluster viruses (Table 2). With the exception of one study, protection was conferred despite low levels of HI cross-reactive antibodies (HI GMT to challenge virus ≤20). For these studies, the threshold associated with protective efficacy was defined as the lowest total EpiCC score (-0.002; MN02 H1β) comparing these five challenge strains with the vaccine strain (IA00 H1γ FS). This threshold defines the white area in Fig. 6. For strains with EpiCC scores above the threshold, the scores of shared epitopes represented at least 66.1% (0.078) of the vaccine's baseline EpiCC score (0.118; Supplementary Table 1), which may suggest that a field strain or a vaccine strain (or both) may have many unique epitopes, but as long as there is a sufficient level of shared epitopes relative to the baseline, a vaccine will be protective.

In addition, the IA00 H1γ vaccine induced partial protection against challenge with IL08 H1α. This EpiCC score (-0.038) was considered a threshold associated with partial protection. This threshold separates the light grey from the dark grey area in Fig. 6. IL08 H1α's shared epitopes score was 57.6% (0.068) of the vaccine's baseline EpiCC score. EpiCC scores for the H1δ cluster viruses, and the IA30 cH1 and SD15 H1α strains were below this threshold in the dark grey area.

Based on the association between total EpiCC scores of HA and vaccine efficacy, we speculate that immunization with the IA00 H1γ FS vaccine strain could confer protection against challenge with viruses with scores above -0.002 (white area in Fig. 6) and partial protection for challenge-vaccine EpiCC scores between -0.002 and -0.038 (light gray area in Fig. 6), assuming minimal variation of internal antigens. In contrast, the vaccine might not protect against viruses for which EpiCC scores fall in the dark gray area of Fig. 6. Important differences between strain identity and EpiCC scores can be identified; for example, the amino acid sequence of IA12 H1γ-2 virus HA is highly identical to IA00 H1γ FS (93.36%); however, its total EpiCC score (-0.023) is below the threshold associated with protection, but above the threshold for partial protection. The low total EpiCC score is driven by a low score for shared epitopes (0.071; 60.2%) and high score for unique epitopes (0.093; Supplemental Table 1). Nevertheless, shared epitopes might still contribute to a certain level of protection.

Discussion

EpiCC is a method for assessing the relationship between field and vaccine strains of pathogens using predicted T cell epitope content as a metric for comparison. As compared to standard methods for estimating vaccine efficacy, such as determining whether immunization induces cross-reactive antibodies to the HA proteins, or measuring genetic differences by sequence similarity, EpiCC characterizes the differences based on portions of the virus that the immune system processes and presents to T cells that drive protective responses. The EpiCC calculation considers the epitope content shared between sequences and penalizes strain and vaccine unique epitopes. As a first test of the EpiCC scoring system, we compared the T cell epitope content of 20 HA sequences from different H1 clusters present in the North American swine population to that of three HA sequences from H1 viruses contained in a commercial swine IAV vaccine. To evaluate whether T cell epitope relatedness between vaccine and non-homologous challenge strains was associated with protection, we compared EpiCC scores with experimental outcomes of efficacy studies of FS H1γ vaccine virus where protection was induced without cross-reactive HA antibodies. The results of the analysis, performed without foreknowledge of efficacy outcomes, showed a threshold of T cell epitope relatedness that explained protection.

We do not yet know whether the threshold score would apply to new strains or different vaccines. The thresholds described in this study were based on experimental data from only six vaccine efficacy studies against challenge with cH1-lineage viruses. To evaluate whether the thresholds could be applied to other vaccines, we compared the epitope content predicted using supertype class I and II HLA alleles of the HA sequence of A/Brisbane/59/2007, the H1 virus in the pre-pandemic (2008-2009) conventional influenza vaccine, and that of CA09 H1pdm. The total EpiCC score (-0.023) fell in the area of partial protection, which is consistent with the reduced influenza-like illnesses and confirmed infection among older adults in the absence of cross-reactive antibodies [34]. However, additional efficacy studies would help to refine the thresholds for prediction of protection and partial protection. Further studies are also required to extend these findings to other IAV strains, and to determine whether EpiCC scores can be used to define thresholds of vaccine efficacy for other economically important pathogens affecting the swine industry.

We note that the set of MHC alleles used for the prediction of epitopes influenced the scores of shared and unique epitopes, and therefore the EpiCC scores, between vaccine strains and field viruses. To illustrate this point, EpiCC scores calculated using binding predictions to SLA alleles were shown to be different from those determined using a set of supertype HLA alleles. For this study, SLA allele selection was based on frequencies determined using low-resolution haplotyping for a small number of pigs [22]. The relevance of these differences using distinct sets of MHC alleles is unknown; however, the distribution of SLA alleles for the North American swine population has yet to be defined, and therefore the EpiCC scores might be different using a more comprehensive set of alleles. Development of a high-throughput SLA typing system paired with a systematic study of SLA diversity would improve the utility of the EpiCC analysis for swine populations not only for IAV, but also for other economically important pathogens affecting the swine industry. Nevertheless, EpiCC scores may be useful for estimating vaccine efficacy for populations of swine for which SLA types are well-defined (e.g. for commercial pork operations where breeding practices limit SLA diversity).

We do not know which component of the score (shared class II epitopes or shared class I epitopes, for example), is more important for predicting protection. Published information describing T cell-dependent (CMI) responses elicited by swine IAV vaccines is scarce and some studies have

reported that CMI responses to inactivated vaccines can be limited in pigs [35]. However, other studies showed that inactivated vaccines can prime the CD4+CD8+ (double-positive) memory T cell subset [9,36,37]. Porcine CD4+CD8+ T cells are MHC class II-restricted memory cells that have T helper function [38]; and also express perforin and mediate cytolytic activity against virus-infected cells [39]. For the set of alleles used for epitope prediction in this analysis, we found higher scores for class II epitopes shared between vaccines and field virus HAs compared to those of class I. Should further studies determine that cross-reactive class II epitopes are more relevant for vaccine protection, a weighted EpiCC score that favors class II epitopes could be applied. Furthermore, alternative versions of the calculation (with or without scores of unique T cell epitopes) were tested. Future studies using more vaccine efficacy data will consider all these versions to determine predictive advantage of one calculation over the others.

EpiCC scores showing high levels of T cell epitope relatedness could explain how protection against challenge can be observed in an absence of HI cross-reactivity in experimental efficacy studies of the FS γ-cluster vaccine virus. The IA00 H1γ FS vaccine virus was genetically and antigenically distinct from the challenge viruses used in the six experimental challenge studies for which efficacy data were available. However, under the conditions of these experimental studies, vaccination provided protection or partial protection against MN11 H1γ, IA92 H1α, OH10 H1γ, and CA09 H1pdm with HI titers lower than 1:40 (the cutoff generally considered predictive of protection) [6,9–11]. Accordingly, MN11 H1γ, IA92 H1α, and OH10 H1γ had the highest EpiCC scores among evaluated HA sequences when compared to IA00 H1γ FS; CA09 H1pdm had the sixth highest score. Among these viruses, only IA92 H1α and CA09 H1pdm have different internal genes [3]. IA92 H1α predates the emergence of TRIG, and CA09 H1pdm was classified as a swine-origin IAV because internal and HA gene segments were genetically similar to those in the triple-reassortant viruses circulating in North American swine [40]. Some differences in strain-specific T cell epitope content of internal proteins should be expected. Notwithstanding these potential differences, the IA00 H1γ FS vaccine was protective against challenge with both strains [6,9]. This result may suggest that a certain level of shared T cell epitopes could be associated with protection, despite the presence of unique epitopes. Antibodies to other surface antigens may have also played a role in protection.

This analysis was limited to T cell epitopes predicted from the sequences of the highly variable external IAV protein, HA. The specific role of HA T cell epitopes in protection against influenza in pigs is not yet known; however, human studies showed that vaccination with a monovalent subunit CA09 H1pdm vaccine elicited robust HA-specific CD4 T cell responses dominated by memory CD4 T cells specific for peptides shared between the seasonal and pandemic strain. Researchers also demonstrated that expansion of CD4 T cells specific for peptide epitopes within HA, but not NP, correlated with neutralizing antibody response [41,42]. These results support the notion that a greater degree of CD4 T cell cross-reactivity may be responsible for the better antibody response.

CMI responses directed to the conserved internal proteins of influenza viruses are highly cross-reactive [43,44]. In pigs, two evolutionary lineages (H1pdm09 and TRIG) dominate the selection of internal genes in the circulating influenza viruses, leading to a high degree of conservation in the internal genes [3,23,24]. As a result, much of the antigenic variability between IAV strains circulating in swine is determined by variable surface antigens (HA, NA and M2). We previously identified SLA-restricted epitopes derived from HA, NA and M conserved in IAV from different HA subtypes [22]. Other groups have reported SLA class I-restricted epitopes in HA and NA [45,46]. Future studies will compare the utility of including other surface antigens (such as NA and M2) in the EpiCC score to determine whether thresholds revealed by HA-specific T cell epitope relatedness can be further refined.

Genetic sequence comparison of HA is also used for predicting potential cross-protective efficacy of vaccines. The relationship between EpiCC scores and HA amino acid sequence identity was nonlinear. At approximately 40% identity, epitope-based relatedness was similar to the EpiCC score of a random amino acid sequence of the same length. And while changes in amino acids affecting T cell epitopes can have a significant impact on the immunogenicity of an antigen [47,48], their effect on whole antigen sequence similarity may be minimal. To illustrate this point, we found viruses that had low EpiCC scores despite having high sequence identity. For example, when IA00 H1γ FS is compared to IA12 H1γ-2 the EpiCC score is above the threshold set for partial protection, but below the threshold for protection. Although H1γ-2 cluster viruses were infrequently detected in the U.S. swine population, characterization of H1γ-2 viruses demonstrated divergent antigenic properties with viruses within the same clade and viruses from contemporary swine H1 clusters as well as commercial vaccines,

suggesting a potential risk of vaccine failure against H1γ-2 viruses [3]. On the contrary, EpiCC analysis suggests that the H1γ cluster vaccine virus may induce at least partial protection against IA12 H1γ-2 virus.

In conclusion, we developed the EpiCC algorithm to assess the immunologic relatedness between antigens based on their predicted T cell epitope content. Using EpiCC, we found that vaccine protection conferred by the FS IA00 H1 γ-cluster, in the absence of HI cross-reactive antibodies, might be explained by predicted T cell epitope content relatedness between challenge and vaccine viruses. Based on these results, we proposed EpiCC score thresholds for prediction of full and partial protection. EpiCC scores were dependent on a set of swine MHC alleles used for the predictions of epitopes; thus future EpiCC scores for these sequences may differ from the scores reported here. As information about SLA prevalence in North American swine populations becomes available, the impact of MHC allele selection on EpiCC scores will be assessed further. Despite these limitations, this study provides preliminary evidence that EpiCC may be of use as an additional measure for selecting the best-matched vaccine virus for immunization against IAV in a herd, and possibly to help predict whether current vaccines would protect against novel viruses introduced into the swine population. Additional data from vaccine efficacy studies will be useful to validate and optimize these thresholds.

Acknowledgments

The authors gratefully acknowledge helpful discussions with Dr. Chris Bailey-Kellogg and Ryan Tassone in the preparation of this manuscript.

Competing interests

Funding for this project was provided by University of Rhode Island and Zoetis Inc. VRG was employed by Zoetis Inc. at the time this study was conducted. ADG and WDM are senior officers and majority shareholders at EpiVax, Inc., a privately-owned immunoinformatics and vaccine design company located in Providence, RI USA. LM and FET are employees at EpiVax, in which LM holds stock options. ADG, WDM, LM and FET acknowledge that there is a potential conflict of interest related to their relationship with EpiVax and attest that the work contained in this research report is free

of any bias that might be associated with the commercial goals of the company. The study was performed using data provided by Zoetis; efficacy outcomes were not considered until after the EpiCC scores were calculated. While Zoetis scientists participated in the analysis of the results, the analysis was performed independently by AHG, ADG and LM, limiting the potential for bias in the outcome of this study.

References

1. Sandbulte MR, Spickler AR, Zaabel PK, Roth JA. Optimal Use of Vaccines for Control of Influenza A Virus in Swine. Vaccines (Basel). 2015;3:22–73.

2. Vincent A, Awada L, Brown I, Chen H, Claes F, Dauphin G, et al. Review of Influenza A Virus in Swine Worldwide: A Call for Increased Surveillance and Research. Zoonoses Public Health. 2014;61:4–17.

3. Anderson TK, Campbell BA, Nelson MI, Lewis NS, Janas-Martindale A, Killian ML, et al. Characterization of co-circulating swine influenza A viruses in North America and the identification of a novel H1 genetic clade with antigenic significance. Virus Res. 2015;201:24–31.

4. Vincent AL, Ma W, Lager KM, Gramer MR, Richt JA, Janke BH. Characterization of a newly emerged genetic cluster of H1N1 and H1N2 swine influenza virus in the United States. Virus Genes. 2009;39:176–85.

5. Lorusso A, Vincent AL, Harland ML, Alt D, Bayles DO, Swenson SL, et al. Genetic and antigenic characterization of H1 influenza viruses from United States swine from 2008. J. Gen. Virol. 2011;92:919–30.

6. Vincent AL, Ciacci-Zanella JR, Lorusso A, Gauger PC, Zanella EL, Kehrli Jr. ME, et al. Efficacy of inactivated swine influenza virus vaccines against the 2009 A/H1N1 influenza virus in pigs. Vaccine. 2010;28:2782–7.

7. Gramer M, Rossow K. Epidemiology of swine influenza and implications of reassortment. Allen D. Leman Swine Conference. 2004. p. 69–73.

8. Rapp-Gabrielson VJ, Wicklund E, Ficken M. Efficacy of FluSure® against challenge with heterologous reassortant swine influenza H1N1 viruses. Allen D. Leman Swine Conference. 2005. p. 14.

9. Kitikoon P, Nilubol D, Erickson BJ, Janke BH, Hoover TC, Sornsen SA, et al. The immune response and maternal antibody interference to a heterologous H1N1 swine influenza virus infection following vaccination. Vet. Immunol. Immunopathol. 2006;112:117–28.

10. Rapp-Gabrielson VJ, Lenz M, Hildebrand T, Taylor L, Kuhn M. Evaluation of cross-protection of FluSure XP® against a heterologous gamma cluster H1N1 swine influenza virus challenge. Allen D. Leman Swine Conference. 2011. p. 267.

11. Rapp-Gabrielson VJ, Lenz M, Kuhn M, Taylor L, Culhane M, Kesl L, et al. Cross-protection of FluSure® XP in pigs challenged with a Gamma cluster H1N1/pH1N1 reassortant swine influenza virus. Allen D. Leman Swine Conference. 2013. p. 219.

12. Detmer SE, Gramer MR, King VL, Mathur S, Rapp-Gabrielson VJ. In vivo evaluation of vaccine efficacy against challenge with a contemporary field isolate from the α cluster of H1N1 swine influenza virus. Can. J. Vet. Res. 2013;77:24–32.

13. La Gruta NL, Turner SJ. T cell mediated immunity to influenza: mechanisms of viral control. Trends Immunol. 2014;35:396–402.

14. Altenburg AF, Rimmelzwaan GF, de Vries RD. Virus-specific T cells as correlate of (cross-)protective immunity against influenza. Vaccine. 2015;33:500–6.

15. Weinfurter JT, Brunner K, Capuano SV, Li C, Broman KW, Kawaoka Y, et al. Cross-reactive T cells are involved in rapid clearance of 2009 pandemic H1N1 influenza virus in nonhuman primates. PLoS Pathog. 2011;7:e1002381.

16. Moise L, Tassone R, Latimer H, Terry F, Levitz L, Haran JP, et al. Immunization with cross-conserved H1N1 influenza CD4+ T-cell epitopes lowers viral burden in HLA DR3 transgenic mice. Hum Vaccin Immunother. 2013;9:2060–8.

17. Price GE, Soboleski MR, Lo C-Y, Misplon JA, Pappas C, Houser KV, et al. Vaccination focusing immunity on conserved antigens protects mice and ferrets against virulent H1N1 and H5N1 influenza A viruses. Vaccine. 2009;27:6512–21.

18. Epstein SL, Price GE. Cross-protective immunity to influenza A viruses. Expert Review of Vaccines. 2010;9:1325–41.

19. Rock KL, Reits E, Neefjes J. Present Yourself! By MHC Class I and MHC Class II Molecules. Trends Immunol. 2016;37:724-37.

20. Gutiérrez AH, Martin WD, Bailey-Kellogg C, Terry F, Moise L, De Groot AS. Development and validation of an epitope prediction tool for swine (PigMatrix) based on the pocket profile method. BMC Bioinformatics. 2015;16:290.

21. Terry FE, Moise L, Martin RF, Torres M, Pilotte N, Williams SA, et al. Time for T? Immunoinformatics addresses vaccine design for neglected tropical and emerging infectious diseases. Expert Rev Vaccines. 2015;14:21–35.

22. Gutiérrez AH, Loving C, Moise L, Terry FE, Brockmeier SL, Hughes HR, et al. In vivo validation of predicted and conserved T cell epitopes in a swine influenza model. PLoS ONE. 2016;11:e0159237.

23. Nelson MI, Detmer SE, Wentworth DE, Tan Y, Schwartzbard A, Halpin RA, et al. Genomic reassortment of influenza A virus in North American swine, 1998-2011. J. Gen. Virol. 2012;93:2584–9.

24. Anderson TK, Nelson MI, Kitikoon P, Swenson SL, Korslund JA, Vincent AL. Population dynamics of cocirculating swine influenza A viruses in the United States from 2009 to 2012. Influenza Other Respir Viruses. 2013;7 Suppl 4:42–51.

25. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 2016;

26. Southwood S, Sidney J, Kondo A, del Guercio MF, Appella E, Hoffman S, et al. Several common HLA-DR types share largely overlapping peptide binding repertoires. J. Immunol. 1998;160:3363–73.

27. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. BMC Immunology. 2008;9:1.

28. De Groot AS, Jesdale BM, Szu E, Schafer JR, Chicz RM, Deocampo G. An interactive Web site providing major histocompatibility ligand predictions: application to HIV research. AIDS Res. Hum. Retroviruses. 1997;13:529–31.

29. Im E-J, Hong JP, Roshorm Y, Bridgeman A, Létourneau S, Liljeström P, et al. Protective efficacy of serially up-ranked subdominant CD8+ T cell epitopes against virus challenges. PLoS Pathog. 2011;7:e1002041.

30. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000;16:276–7.

31. Shen ZT, Nguyen TT, Daniels KA, Welsh RM, Stern LJ. Disparate epitopes mediating protective heterologous immunity to unrelated viruses share peptide-MHC structural features recognized by cross-reactive T cells. J. Immunol. 2013;191:5139–52.

32. Moise L, Gutierrez AH, Bailey-Kellogg C, Terry F, Leng Q, Abdel Hady KM, et al. The two-faced T cell epitope: Examining the host-microbe interface with JanusMatrix. Hum Vaccin Immunother. 2013;9:1577–86.

33. Rudolph MG, Stanfield RL, Wilson IA. How Tcrs Bind Mhcs, Peptides, and Coreceptors. Annual Review of Immunology. 2006;24:419–66.

34. Jhung MA, Swerdlow D, Olsen SJ, Jernigan D, Biggerstaff M, Kamimoto L, et al. Epidemiology of 2009 pandemic influenza A (H1N1) in the United States. Clin. Infect. Dis. 2011;52 Suppl 1:S13-26.

35. Loving CL, Vincent AL, Pena L, Perez DR. Heightened adaptive immune responses following vaccination with a temperature-sensitive, live-attenuated influenza virus compared to adjuvanted, whole-inactivated virus in pigs. Vaccine. 2012;30:5830–8.

36. Platt R, Vincent AL, Gauger PC, Loving CL, Zanella EL, Lager KM, et al. Comparison of humoral and cellular immune responses to inactivated swine influenza virus vaccine in weaned pigs. Vet. Immunol. Immunopathol. 2011;142:252–7.

37. Sandbulte MR, Platt R, Roth JA, Henningson JN, Gibson KA, Rajão DS, et al. Divergent immune responses and disease outcomes in piglets immunized with inactivated and attenuated H3N2 swine influenza vaccines in the presence of maternally-derived antibodies. Virology. 2014;464–465:45–54.

38. Gerner W, Käser T, Saalmüller A. Porcine T lymphocytes and NK cells – An update. Developmental & Comparative Immunology. 2009;33:310–20.

39. Denyer MS, Wileman TE, Stirling CMA, Zuber B, Takamatsu H-H. Perforin expression can define CD8 positive lymphocyte subsets in pigs allowing phenotypic and functional analysis of natural killer, cytotoxic T, natural killer T and MHC un-restricted cytotoxic T-cells. Vet. Immunol. Immunopathol. 2006;110:279–92.

40. Smith GJD, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature. 2009;459:1122–5.

41. Alam S, Knowlden ZAG, Sangster MY, Sant AJ. CD4 T cell help is limiting and selective during the primary B cell response to influenza virus infection. J. Virol. 2014;88:314–24.

42. DiPiazza A, Richards KA, Knowlden ZAG, Nayak JL, Sant AJ. The Role of CD4 T Cell Memory in Generating Protective Immunity to Novel and Potentially Pandemic Strains of Influenza. Front Immunol. 2016;7:10.

43. Assarsson E, Bui H-H, Sidney J, Zhang Q, Glenn J, Oseroff C, et al. Immunomic analysis of the repertoire of T-cell specificities for influenza A virus in humans. J. Virol. 2008;82:12241–51.

44. Lee LY-H, Ha DLA, Simmons C, de Jong MD, Chau NVV, Schumacher R, et al. Memory T cells established by seasonal human influenza A infection cross-react with avian influenza A (H5N1) in healthy individuals. J. Clin. Invest. 2008;118:3478–90.

45. Zhang N, Qi J, Feng S, Gao F, Liu J, Pan X, et al. Crystal structure of swine major histocompatibility complex class I SLA-1 0401 and identification of 2009 pandemic swine-origin influenza A H1N1 virus cytotoxic T lymphocyte epitope peptides. J. Virol. 2011;85:11709–24.

46. Pedersen LE, Breum SO, Riber U, Larsen LE, Jungersen G. Identification of swine influenza virus epitopes and analysis of multiple specificities expressed by cytotoxic T cell subsets. Virol. J. 2014;11:163.

47. Moise L, Song C, Martin WD, Tassone R, De Groot AS, Scott DW. Effect of HLA DR epitope de-immunization of Factor VIII in vitro and in vivo. Clin. Immunol. 2012;142:320–31.

48. Zhao H, Verma D, Li W, Choi Y, Ndong C, Fiering SN, et al. Depletion of T cell epitopes in lysostaphin mitigates anti-drug antibody response and enhances antibacterial efficacy in vivo. Chem. Biol. 2015;22:629–39.

Table 1. HA sequence information for swine H1 IAV.

Virus name[a]	Virus H1 cluster	Label[b]	GenBank accession or source[c]
A/swine/Iowa/15/1930 (H1N1)	classical	IA30 cH1	EU139823
A/swine/Illinois/02450/2008 (H1N1)	α	IL08 H1α	CY099052
A/swine/South Dakota/A01823598/2015	α	SD15 H1α	KT356682
A/swine/St-Hyacinthe/106/1991 (H1N1)	α	SH91 H1α	U11857
A/swine/Iowa/40766/1992 (H1N1)	α	IA92 H1α	KP788773
A/swine/Minnesota/00040/2002 (H1N1)	β	MN02 H1β	Zoetis
A/swine/Iowa/00239/2004 (H1N1)	β	IA04 H1β	KM198690
A/swine/Iowa/110600/2000 (H1N1)	γ	IA00 H1γ FS	Zoetis
A/swine/Minnesota/PAH618/2011 (H1N1)	γ	MN11 H1γ	Zoetis
A/swine/Ohio/02973/2010 (H1N1)	γ	OH10 H1γ	Zoetis
A/swine/Iowa/A01940123/2015 (H1N1)	γ	IA15 H1γ	KT699044
A/swine/Minnesota/A01940015/2015 (H1N1)	γ	MN15 H1γ	KT595733
A/swine/Iowa/A01410129/2012 (H1N1)	γ2	IA12 H1γ-2	KJ397936
A/California/04/2009 (H1N1)	H1N1pdm09	CA09 H1pdm	GQ117044
A/swine/Oklahoma/0726H/2008 (H1N2)	δ1	OK08 H1δ1 FS	Zoetis
A/swine/Ontario/55383/04 (H1N2)	δ1	ON04 H1δ1	DQ280212
A/swine/Illinois/PAH710/2011 (H1N2)	δ1	IL11 H1δ1	Zoetis
A/swine/South Dakota/A01823304/2015	δ1	SD15 H1δ1	KT277819
A/swine/Oklahoma/A01566774/2014 (H1N2)	δ1	OK14 H1δ1	KP270784
A/swine/Minnesota/A01823864/2015 (H1N2)	δ1	MN15a H1δ1	KT699050
A/swine/Iowa/A01823426/2015 (H1N2)	δ1	IA15 H1δ1	KT356694
A/swine/Minnesota/A01940042/2015 (H1N2)	δ1	MN15b H1δ1	KT733589
A/swine/North Carolina/031/2005 (H1N1)	δ2	NC05 H1δ2 FS	Zoetis
A/swine/NC/00573/2005 (H1N1)	δ2	NC05 H1δ2	FJ638306

[a]FS viruses are shown in bold font.
[b]FS viruses have 'FS' at the end of their labels.
[c]Sequences marked ‘Zoetis’ were provided by Zoetis and are considered proprietary.

Table 2. FluSure XP® vaccination and challenge studies.

Heterologous challenge	Measurement of protection[a]			HI GMT to challenge virus	Outcome	Ref.
Virus	Percentage of macroscopic pneumonia	Virus titers in nasal swabs	Virus titers in lungs			
MN02 H1β	Reduced	Reduced	Not available	80	Protection	7
IA92 H1α	Significantly reduced	Significantly reduced	Not available	≤20	Protection	9
CA09 H1pdm	Significantly reduced	Significantly reduced	Significantly reduced	≤10	Protection	6
OH10 H1γ	Significantly reduced	Not available	Significantly reduced	≤10	Protection	10
MN11 H1γ	Significantly reduced	Significantly reduced	Significantly reduced	≤20	Protection	11
IL08 H1α	Not significantly different	Significantly reduced	Significantly reduced	≤20	Partial protection	12

[a]Significance of outcomes was as measured and reported in the original references.

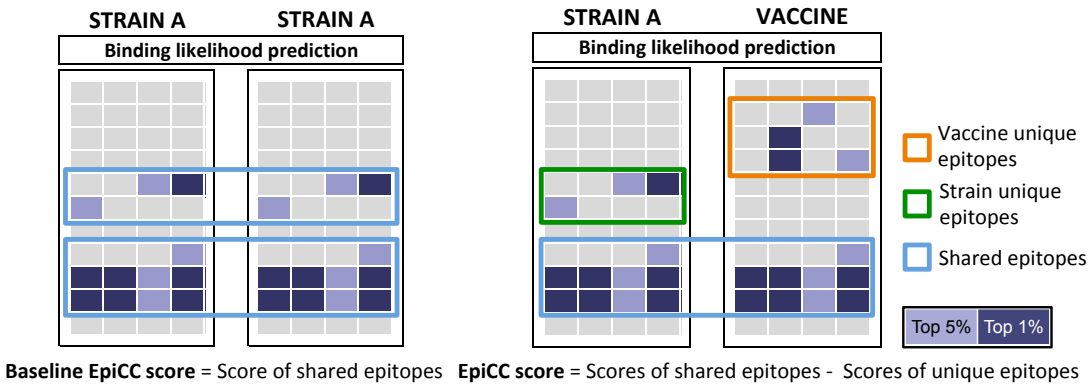


Fig. 1. Illustration of T cell epitope content comparison (EpiCC) score calculation. Strain A and Vaccine were screened for binding likelihood to a set of four MHC alleles; 9-mers (rows) predicted to bind to specific MHC alleles (columns) are shown in light (top 5%) or dark (top 1%) blue. The comparison of the epitope content of Strain A to itself determines the baseline EpiCC score (left). For the comparison between Strain A and Vaccine, scores of shared and unique epitopes are considered when calculating the EpiCC score (right).

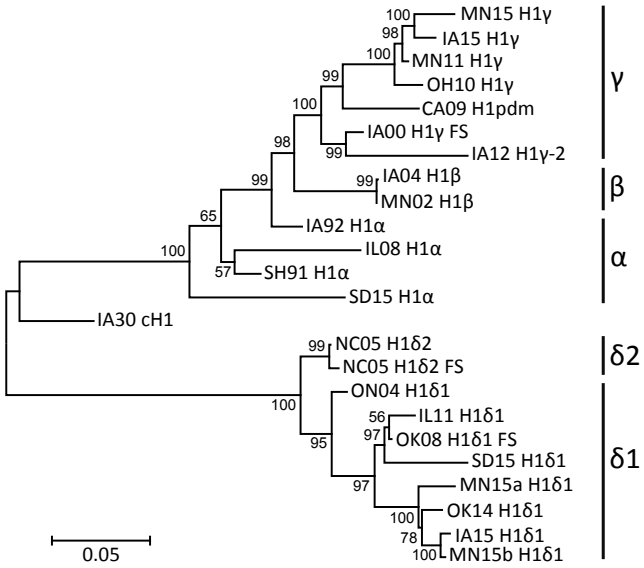


Fig. 2. Phylogenetic tree for the HA nucleotide sequences of IAV field and vaccine viruses representing H1 phylo-cluster in the North American swine. Bootstrap test results are shown next to the branches.

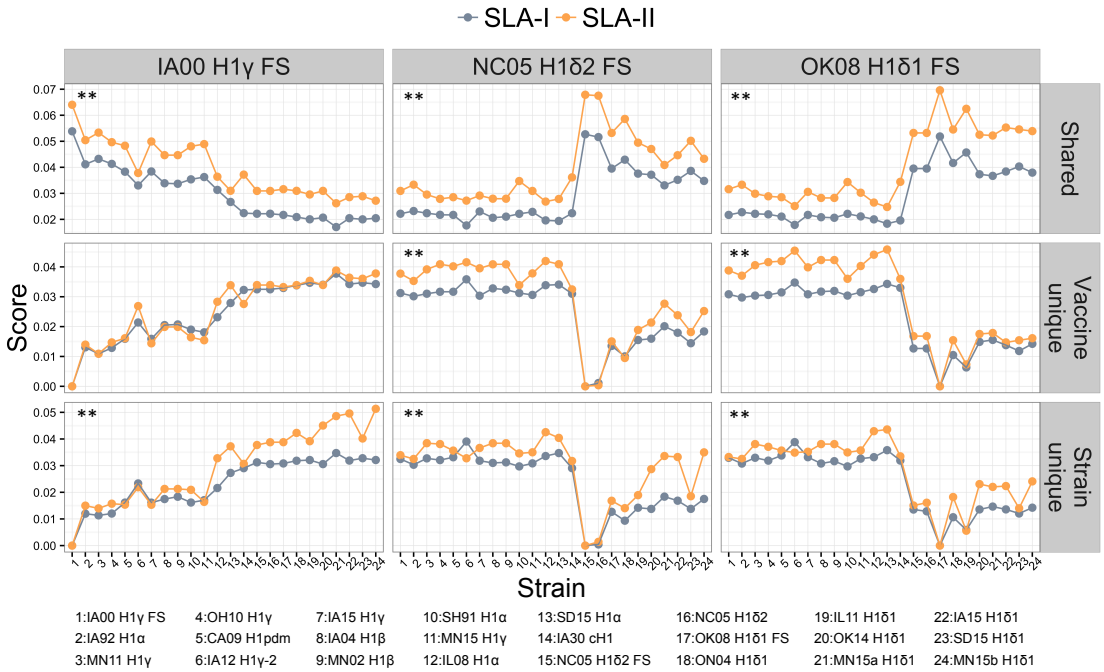


Fig. 3. Comparison of scores of shared and unique epitopes across strains. Scores of shared, vaccine unique and strain unique SLA class I and II epitopes were determined for the comparison of HA sequences from vaccine viruses and field (challenge) strains. Note that y-axes show different scales. Solid connecting lines are included only for visualization purposes. P-values of comparisons were calculated using one-tailed Wilcoxon matched-pairs signed rank test (**$p<0.001$). HA vaccine sequences had higher scores for shared epitopes with strains belonging to the same H1 cluster or the same HA lineage. In general, scores of class II shared and unique epitopes were significantly higher than those of class I. Viruses are sorted by nucleotide identity relative to H1γ FS. Strain numbers on the x-axis are described in detail in the legend below.

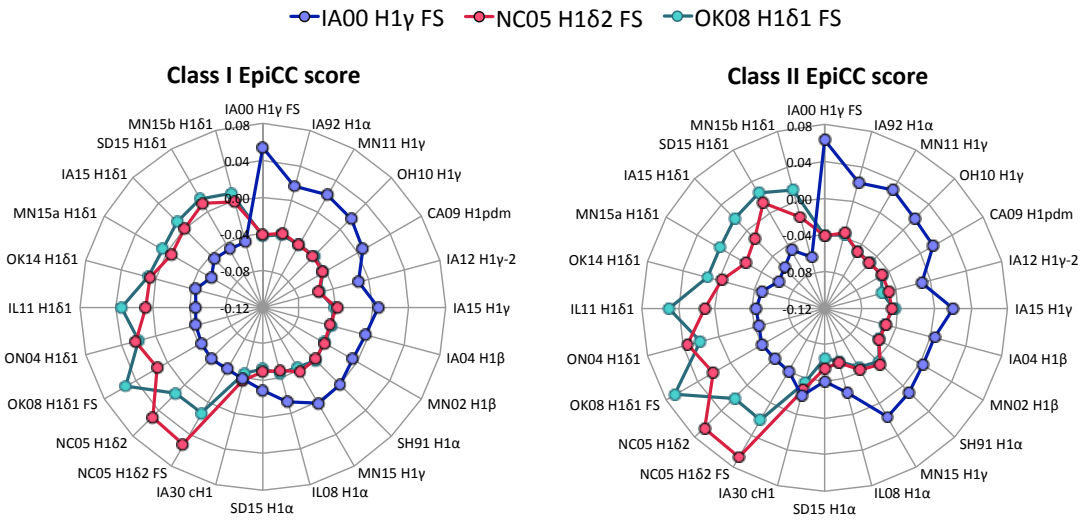


Fig. 4. EpiCC score comparisons between HA sequences of FS vaccine viruses and field viruses. Each axis corresponds to the HA sequence of one virus. EpiCC score = Scores of shared epitopes - Scores of strain and vaccine unique epitopes.

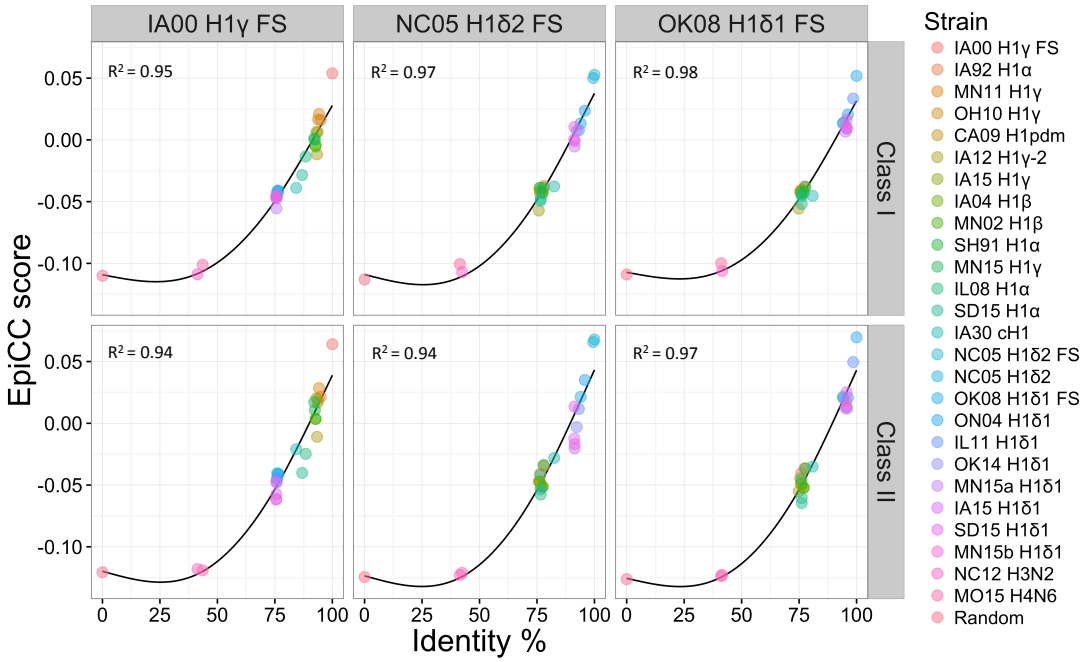


Fig. 5. Relationship between EpiCC scores and identity. The second order polynomial relationship between class I (top), II (bottom) EpiCC scores, and amino acid identity for each FS virus are shown. R^2 of regression models are shown. H3N2, H4N6 and a random sequence were included in this analysis to represent the lower end of the identity range. Interestingly, there were instances where viruses had low EpiCC scores despite high identity (e.g. class II epitope content of IA00 H1γ FS compared to SD15 H1α).

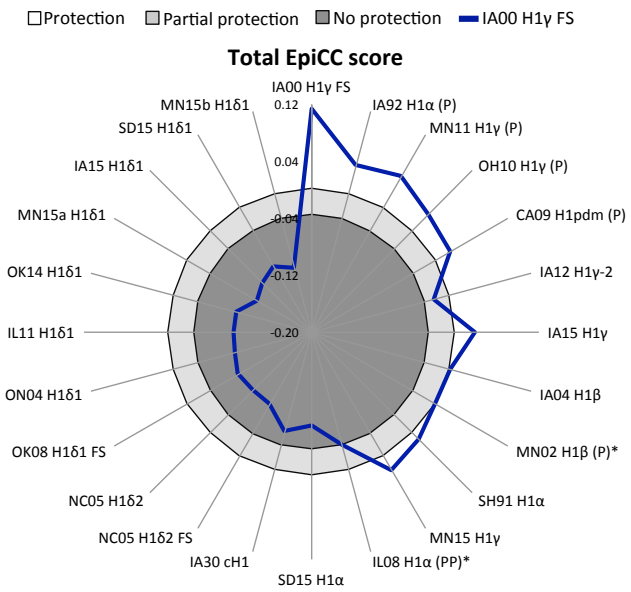


Fig. 6. Definition of threshold for prediction of vaccine efficacy prediction. Total EpiCC scores (blue line) for the comparison of H1γ FS and each viral strain are shown. The FS γ-cluster vaccine strain was protective or partially protective against challenge with viruses annotated as (P) or (PP), respectively. The rest of the viruses were not tested as challenge strains. Protection and partial protection thresholds (black lines) defined three areas shown in white (protection; total EpiCC score above -0.002), light gray (partial protection) and dark gray (no protection). Viruses used to set the thresholds are marked with an asterisk (*). We hypothesize that FS would confer at least partial protection against challenge with viruses that had EpiCC scores outside the darker gray region.

CHAPTER 5

Conclusions

The development of immunoinformatics tools has had a significant impact on human immunology research, discovery of immunogenic epitopes and design of vaccines for diverse pathogens. One of the main applications of immunoinformatics is T cell epitope prediction due to their critical role in cell-mediated immunity (CMI). T cell epitope mapping algorithms can easily analyze complete proteomes of any size to identify candidates for further experimental validation. However, immunoinformatics tools are very limited for non-human species. In this thesis, concepts and information used to construct immunoinformatics methods for humans were applied to develop novel T cell epitope-based immunoinformatics tools for swine.

PigMatrix has been designed to map swine T cell epitopes using the pocket profile method. This method was applied to overcome the lack of SLA-specific binding data required to train and test predictors by leveraging predefined binding preferences for HLA. The pocket profile method has been previously applied for prediction of binding to HLA alleles with limited binding data [1,2]. This is the first formal application and evaluation of the method for a non-human species. Different approaches were tested to define pocket residues (based on SLA or HLA structures) and similarity between SLA and HLA pockets for selection of the best human match. Without any training step, PigMatrix achieved a favorable predictive performance, comparable to or better than other available methods for SLA class I alleles. It is noteworthy that PigMatrix is the first tool designed for prediction of SLA class II T cell epitopes.

The current implementation of PigMatrix compares the sequences of an SLA pocket to a library of HLA pockets and calculates a similarity score to identify the best human match and infer binding preferences. The library of HLA pockets is comprised of HLA supertype alleles, but it can be readily extended to include more alleles with defined binding specificities. Thus, it would be possible to find better human matches, which may improve PigMatrix predictions. Furthermore, weighting the similarity score by conservation of key contact residues or biochemical properties in the pocket may also improve the selection of the best human match.

Larger datasets of SLA-restricted peptides are required to further evaluate the PigMatrix

approach and improve the predictions. To generate quantitative binding data and test PigMatrix, binding assays for commonly expressed SLA molecules could be developed. These assays would provide valuable information to better define binding preferences and potentially develop predictions based on SLA specificities rather than HLA pocket preferences. Currently, binding assays are limited to only three SLA class I alleles [3]. In addition, crystallographic structures of diverse SLA molecules would further expand our understanding of the SLA binding pockets and their specificities; only two crystal structures of SLA class I alleles are currently available [4,5]. Similar to class I predictions, class II predictions may improve if pocket residues are defined based on SLA crystal structures rather than structures of HLA molecules.

Based on the initial encouraging results of PigMatrix, predictions were extended to SLA alleles reported to be prevalent in commercial pigs and applied to identify potentially immunogenic IAV peptides. IAV is considered one of the most important infectious disease agents affecting North American swine [6,7]. The majority of currently licensed swine IAV vaccines consist of whole inactivated viruses administered with adjuvant by intramuscular injection [8]. This platform primarily induces systemic IgG antibody responses to the surface glycoproteins, mainly HA [8,9]. However, antibody-mediated immunity does not typically provide protection against divergent strains of IAV [9,10]. Conversely, CMI can be broadly cross-reactive to a variety of IAV subtypes [11–15]. Moreover, CMI contributes to clearance of virus and reduced symptom severity and virus shedding [16,17]. A vaccine that induces CMI and reduces morbidity could prevent anorexia and weight loss, which cause significant economic impact to pork producers. Therefore, identification T cell epitopes conserved in diverse IAV represents the first step toward the development of a potentially broadly protective vaccine. So, PigMatrix was used to screen complete proteomes of representative IAV strains in U.S. swine for class I and II T cell epitopes. A prototype epitope-based DNA vaccine encoding strings of class I and II epitopes was developed to evaluate immunogenic responses to the putative epitopes. Eleven T cell epitopes induced specific IFNγ recall responses in pigs vaccinated with the prototype vaccine. Recall responses to peptides were not observed in pigs immunized with the inactivated commercial vaccine, despite containing similar internal antigens. This result shows that the epitope-based vaccine promoted more efficient processing and presentation of its own epitopes as compared to whole-protein-based

vaccines.

In this immunogenicity study, SLA-typing performed at the end of the study revealed that the SLA alleles used for the prediction were not prevalent in the study cohort, which highlights that the lack of information on SLA frequencies in the U.S. population represents a challenge for the development of T cell epitope predictors for swine. Cohort-specific predictions developed to re-evaluate the putative epitopes showed lower binding likelihood for 46 out of 48 epitopes and were particularly effective identifying non-immunogenic peptides. PigMatrix might be applied to identify T cell epitopes from other swine pathogens, which could significantly speed the development of novel vaccines for infectious diseases in pigs.

Further studies would apply PigMatrix using a more comprehensive set of SLA alleles to identify and add peptides to the set of highly conserved and immunogenic IAV sequences to be included in an epitope-driven vaccine. In addition to SLA binding assays and standard recall assays using blood from naturally infected or vaccinated pigs, the selection process of peptides can be optimized by implementing T cell assays currently applied to evaluate peptide immunogenicity for human studies using naive PBMCs [18]. In this protocol, before performing ELISpot assays, PBMCs are cultured with individual peptides over 8 days to expand peptide-specific T cells ex vivo. Results of this method, in a preliminary evaluation using naïve blood from pigs, showed that it is possible to measure specific IFNγ recall responses upon restimulation with peptides (unpublished).

The prototype vaccine tested, developed as a tool to measure peptide immunogenicity, was designed to target the putative class I and class II epitopes to the endogenous and exogenous antigen presentation pathways, respectively. In this study, pigs were vaccinated three times intramuscularly, but future vaccine efficacy studies could test other immunization methods like gene gun or electroporation. These methods could enhance the *in vivo* transfection (i.e. introduction of nucleic acids into eukaryotic cells) efficiency of the DNA vaccine [19]. To date, three DNA vaccines have been licensed for veterinary use (West Nile Virus vaccine for horses [20], infectious hematopoietic necrosis virus vaccine for salmon [21], and a canine melanoma vaccine [22]). Moreover, the first licensed electroporation-delivered product and first licensed gene therapy was the growth hormone-releasing hormone for use in swine [23]. The development of these products illustrates the advancements in DNA plasmid

technology and its application for animals both big and small. Other methods could be also evaluated to deliver immunogenic T cell epitopes. For example, alphavirus-like replicon particle (RP) technology has recently been approved by the USDA as a vaccine for swine IAV. This vaccine uses RP to deliver a sequence that encodes an H3 HA [8]. RP can be engineered to deliver peptides and an HA component to potentially induce both CMI and humoral responses.

Information about SLA allele diversity in the U.S. swine population is critically important to develop a more comprehensive set of predictions that target the most prevalent SLA alleles. A commercially available high-throughput method for high-resolution SLA-typing would improve the ability of researchers and producers to determine SLA diversity. Researchers in other countries have recognized the importance of SLA diversity for vaccine development and studies to identify commonly expressed haplotypes have been conducted [24]. Once the prevalence and diversity of U.S. swine SLA are better understood, it may be possible to cluster SLA molecules into supertypes. The concept of supertypes has been applied to HLA for selection of few representative alleles from different clusters to cover a high percentage of the HLA diversity in human population [25,26]. An epitope-based vaccine based on peptides predicted to bind SLA supertype alleles could induce immune responses in pigs expressing diverse alleles.

A novel tool for epitope content comparison called EpiCC was developed leveraging PigMatrix predictions. This method estimates the relationship between variable viruses based on their putative T cell epitope content to predict cross-protection. For influenza and other viruses, sequence data and antibody cross-reactivity are commonly used to predict vaccine-induced protection. However, previous efficacy studies demonstrated that in the absence of cross-reactive antibodies, a commercial swine IAV vaccine was capable of inducing protection or partial protection against heterologous challenge strains [9,27–31]. The EpiCC analysis showed that T cell epitope-based relatedness between HA proteins of the vaccine and challenge strains was associated with vaccine-induced protection. The results also demonstrated that the T cell epitope content depends on the set of MHC alleles used for the predictions of epitopes. EpiCC analysis may complement current methods for vaccine selection in outbreak situations and strain selection for vaccine production.

Future studies of EpiCC could explore different criteria to define shared and unique epitopes.

The current definition of shared (or cross-reactive) and unique epitopes is very stringent (only epitopes predicted to bind to the same SLA allele with identical TCR-facing residues were considered cross-reactive); it is likely that this approach underestimates the full set of shared epitopes between vaccines and viruses, but it provides a reasonably conservative means to determine the minimum shared T cell epitope content. In reality, epitope cross-reactivity is conformational, hence, it is possible that epitopes with different TCR-facing residues and completely different sequences can be recognized by the same TCR [32,33]. EpiCC will be tested applying less restrictive criteria for cross-reactivity allowing for conservative mismatches between TCR-facing residues or applying a threshold of similarity [34]. In addition, the initial model assumed equal importance for vaccine and strain unique epitopes as well as for class I and II T cell epitopes; further EpiCC analyses will evaluate the impact of incorporating weighted scores.

The EpiCC analysis was restricted to HA sequences from 23 viruses representing diverse clusters of field strains; we assumed limited T cell epitope variation of other antigens. However, the same approach can be applied to multiple antigens or to complete proteomes of influenza strains or other pathogens. Future studies of large-scale surveillance data will be performed to identify circulating or novel viruses distantly related to current vaccines for further experimental evaluation to determine potential risk of vaccine failure. It would also be possible to identify viruses with low T cell epitope relatedness with vaccine viruses despite high sequence similarity (or vice versa); challenge studies with these viruses will provide evidence of the role of T cell epitopes in protection. Predicted T cell epitopes will be synthesized and tested individually using PBMCs from vaccinated pigs pre- and post-challenge to evaluate PigMatrix predictions as well as responses to shared and unique epitopes. In addition, more sophisticated methods will be explored to visualize the T cell epitope content and the epitope-based relationship among large number of viruses. For example, in preliminary studies, we have tested multidimensional scaling and distance trees to represent T cell epitope-based distances (unpublished). Moreover, EpiCC analysis has obvious applications to human vaccine studies, where epitope content could be compared for predictions using HLA alleles.

The preliminary success of PigMatrix and EpiCC opens the door for the development of other immunoinformatics tools to study the immunogenicity of swine pathogens and to identify potential

regulatory T cell epitopes. Previous large computational studies comparing predicted human T cell epitopes have demonstrated differentiable patterns of TCR cross-reactivity with self, associated with different T cell phenotypes [35,36]. JanusMatrix is the tool developed to perform this analysis; it uses EpiMatrix for prediction of human T cell epitopes and then compares their TCR-facing residues against T cell epitopes predicted within the human proteome, human pathogens, and human microbiome to identify epitopes with identical TCR-facing residues [35]. JanusMatrix has shown that regulatory T cell epitopes have greater TCR cross-reactivity with human proteins than effector T cell epitopes [35]. Integrating both MHC binding and TCR cross-reactivity to self, the Janus Immunogenicity Score has been developed to predict viral and bacterial immunogenicity potential [36,37]. These analyses have provided evidence of a pathogen mechanism of 'immune camouflage' to avoid immune recognition by reducing the MHC and/or TCR binding of their peptides. We are now in the unique position to extend these approaches to swine. Prediction of potential regulatory T cell epitopes may have a significant impact in swine vaccine design. For example, porcine reproductive and respiratory syndrome virus (PRRSV) causes one of the most devastating swine diseases worldwide. One of the mechanisms associated with the ability of PRRSV to negatively modulate the host immune system is the induction of regulatory T cells [38]. A version of JanusMatrix equipped to predict swine T cell epitopes could help to find potential regulatory T cell PRRSV epitopes for experimental evaluation. Thus, PigMatrix and JanusMatrix could help to better classify pathogen epitopes, which will improve future vaccine designs.

Animal researchers are starting to recognize the value and diverse applications of immunoinformatics in the veterinary field. The development of predictive tools for swine T cell epitopes, their application to the analysis of existing vaccines, and the design of new vaccine prototypes, demonstrates the potential of immunoinformatics to aid and accelerate swine vaccine development and paves the way for future advancements for other important livestock species.

References

1. Zhang H, Lund O, Nielsen M. The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. Bioinformatics. 2009;25:1293–9.

2. Zhang L, Chen Y, Wong H-S, Zhou S, Mamitsuka H, Zhu S. TEPITOPEpan: Extending TEPITOPE for Peptide Binding Prediction Covering over 700 HLA-DR Molecules. PLoS ONE. 2012;7:e30483.

3. Pedersen LE, Jungersen G, Buus S, Golde WT. Analysis of Swine Leukocyte Antigen peptide binding profiles and the identification of T cell epitopes by tetramer staining [Ph.D. thesis]. Technical University of Denmark; 2012.

4. Zhang N, Qi J, Feng S, Gao F, Liu J, Pan X, et al. Crystal structure of swine major histocompatibility complex class I SLA-1 0401 and identification of 2009 pandemic swine-origin influenza A H1N1 virus cytotoxic T lymphocyte epitope peptides. J. Virol. 2011;85:11709–24.

5. Fan S, Wu Y, Wang S, Wang Z, Jiang B, Liu Y, et al. Structural and Biochemical Analyses of Swine Major Histocompatibility Complex Class I Complexes and Prediction of the Epitope Map of Important Influenza A Virus Strains. J. Virol. 2016;90:6625–41.

6. Donovan TS. The role of influenza on growing pig performance. Allen D. Leman Swine Conference. Saint Paul, MN; 2005. p. 97–8.

7. Wideman G. Impact of Influenza A in Pork Production. Proceedings of the London Swine Conference. London, Ontario; 2013. p. 191–3.

8. Sandbulte MR, Spickler AR, Zaabel PK, Roth JA. Optimal Use of Vaccines for Control of Influenza A Virus in Swine. Vaccines (Basel). 2015;3:22–73.

9. Vincent AL, Ciacci-Zanella JR, Lorusso A, Gauger PC, Zanella EL, Kehrli Jr. ME, et al. Efficacy of inactivated swine influenza virus vaccines against the 2009 A/H1N1 influenza virus in pigs. Vaccine. 2010;28:2782–7.

10. Brandenburg B, Koudstaal W, Goudsmit J, Klaren V, Tang C, Bujny MV, et al. Mechanisms of hemagglutinin targeted influenza virus neutralization. PLoS ONE. 2013;8:e80034.

11. Lee LY-H, Ha DLA, Simmons C, de Jong MD, Chau NVV, Schumacher R, et al. Memory T cells established by seasonal human influenza A infection cross-react with avian influenza A (H5N1) in healthy individuals. J. Clin. Invest. 2008;118:3478–90.

12. Price GE, Soboleski MR, Lo C-Y, Misplon JA, Pappas C, Houser KV, et al. Vaccination focusing immunity on conserved antigens protects mice and ferrets against virulent H1N1 and H5N1 influenza A viruses. Vaccine. 2009;27:6512–21.

13. Babon JAB, Cruz J, Orphin L, Pazoles P, Co MDT, Ennis FA, et al. Genome-wide screening of human T-cell epitopes in influenza A virus reveals a broad spectrum of CD4(+) T-cell responses to internal proteins, hemagglutinins, and neuraminidases. Hum. Immunol. 2009;70:711–21.

14. La Gruta NL, Turner SJ. T cell mediated immunity to influenza: mechanisms of viral control. Trends Immunol. 2014;35:396–402.

15. Altenburg AF, Rimmelzwaan GF, de Vries RD. Virus-specific T cells as correlate of (cross-)protective immunity against influenza. Vaccine. 2015;33:500–6.

16. Lin YL, Askonas BA. Biological properties of an influenza A virus-specific killer T cell clone. Inhibition of virus replication in vivo and induction of delayed-type hypersensitivity reactions. J. Exp. Med. 1981;154:225–34.

17. McMichael AJ, Gotch FM, Noble GR, Beare PA. Cytotoxic T-cell immunity to influenza. N. Engl. J. Med. 1983;309:13–7.

18. Liu R, Moise L, Tassone R, Gutierrez AH, Terry FE, Sangare K, et al. H7N9 T-cell epitopes that mimic human sequences are less immunogenic and may induce Treg-mediated tolerance. Hum Vaccin

Immunother. 2015;11:2241–52.

19. Saade F, Petrovsky N. Technologies for enhanced efficacy of DNA vaccines. Expert Rev Vaccines. 2012;11:189–209.

20. Davis BS, Chang G-JJ, Cropp B, Roehrig JT, Martin DA, Mitchell CJ, et al. West Nile Virus Recombinant DNA Vaccine Protects Mouse and Horse from Virus Challenge and Expresses In Vitro a Noninfectious Recombinant Antigen That Can Be Used in Enzyme-Linked Immunosorbent Assays. J Virol. 2001;75:4040–7.

21. Garver KA, LaPatra SE, Kurath G. Efficacy of an infectious hematopoietic necrosis (IHN) virus DNA vaccine in Chinook Oncorhynchus tshawytscha and sockeye O. nerka salmon. Dis. Aquat. Org. 2005;64:13–22.

22. Bergman PJ, Camps-Palau MA, McKnight JA, Leibman NF, Craft DM, Leung C, et al. Development of a xenogeneic DNA vaccine program for canine malignant melanoma at the Animal Medical Center. Vaccine. 2006;24:4582–5.

23. Draghia-Akli R, Ellis KM, Hill L-A, Malone PB, Fiorotto ML. High-efficiency growth hormone-releasing hormone plasmid vector administration into skeletal muscle mediated by electroporation in pigs. FASEB J. 2003;17:526–8.

24. Pedersen LE, Jungersen G, Sorensen MR, Ho C-S, Vadekær DF. Swine Leukocyte Antigen (SLA) class I allele typing of Danish swine herds and identification of commonly occurring haplotypes using sequence specific low and high resolution primers. Vet. Immunol. Immunopathol. 2014;162:108–16.

25. Southwood S, Sidney J, Kondo A, del Guercio MF, Appella E, Hoffman S, et al. Several common HLA-DR types share largely overlapping peptide binding repertoires. J. Immunol. 1998;160:3363–73.

26. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. BMC Immunology. 2008;9:1.

27. Kitikoon P, Nilubol D, Erickson BJ, Janke BH, Hoover TC, Sornsen SA, et al. The immune response and maternal antibody interference to a heterologous H1N1 swine influenza virus infection following vaccination. Vet. Immunol. Immunopathol. 2006;112:117–28.

28. Gramer M, Rossow K. Epidemiology of swine influenza and implications of reassortment. Allen D. Leman Swine Conference. 2004. p. 69–73.

29. Rapp-Gabrielson VJ, Lenz M, Hildebrand T, Taylor L, Kuhn M. Evaluation of cross-protection of FluSure XP® against a heterologous gamma cluster H1N1 swine influenza virus challenge. Allen D. Leman Swine Conference. 2011. p. 267.

30. Rapp-Gabrielson VJ, Lenz M, Kuhn M, Taylor L, Culhane M, Kesl L, et al. Cross-protection of FluSure® XP in pigs challenged with a Gamma cluster H1N1/pH1N1 reassortant swine influenza virus. Allen D. Leman Swine Conference. 2013. p. 219.

31. Detmer SE, Gramer MR, King VL, Mathur S, Rapp-Gabrielson VJ. In vivo evaluation of vaccine efficacy against challenge with a contemporary field isolate from the α cluster of H1N1 swine influenza virus. Can. J. Vet. Res. 2013;77:24–32.

32. Shen ZT, Nguyen TT, Daniels KA, Welsh RM, Stern LJ. Disparate epitopes mediating protective heterologous immunity to unrelated viruses share peptide-MHC structural features recognized by cross-reactive T cells. J. Immunol. 2013;191:5139–52.

33. Gil A, Kenney LL, Mishra R, Watkin LB, Aslan N, Selin LK. Vaccination and heterologous immunity: educating the immune system. Trans R Soc Trop Med Hyg. 2015;109:62–9.

34. Calis JJA, de Boer RJ, Keşmir C. Degenerate T-cell Recognition of Peptides on MHC Molecules Creates Large Holes in the T-cell Repertoire. PLoS Comput Biol. 2012;8:e1002412.

35. Moise L, Gutierrez AH, Bailey-Kellogg C, Terry F, Leng Q, Abdel Hady KM, et al. The two-faced T cell epitope: Examining the host-microbe interface with JanusMatrix. Hum Vaccin Immunother. 2013;9:1577–86.

36. He L, De Groot AS, Gutierrez AH, Martin WD, Moise L, Bailey-Kellogg C. Integrated assessment of predicted MHC binding and cross-conservation with self reveals patterns of viral camouflage. BMC Bioinformatics. 2014;15:S1.

37. He L, De Groot AS, Bailey-Kellogg C. Hit-and-run, hit-and-stay, and commensal bacteria present different peptide content when viewed from the perspective of the T cell. Vaccine. 2015;33:6922–9.

38. Wongyanin P, Buranapraditkul S, Yoo D, Thanawongnuwech R, Roth JA, Suradhat S. Role of porcine reproductive and respiratory syndrome virus nucleocapsid protein in induction of interleukin-10 and regulatory T-lymphocytes (Treg). J. Gen. Virol. 2012;93:1236–46.

APPENDICES

Supplemental information for Chapter 2

Supplemental Table 1. Peptide database.

Exp: Experimental results; 0: Non-binders, 1: Binder

SLA-1*0401		SLA-2*0401		SLA-3*0401		SLA-DRB1*0201	
Sequence	Exp	Sequence	Exp	Sequence	Exp	Sequence	Exp
ALTDLGLIY	1	AAKHMSNTY	0	AMYDPQTYY	0	AAAPSTTTALDTTPN	0
AMYDPQTYY	1	ASYQFQLPY	1	ASYAAAAAY	0	AAFVTNSTVADELGR	0
ASAAHLAAY	1	ATAAATEAY	1	FLYPSWSLY	0	ADALAPVVVEGERAT	0
ASFAAQLFY	0	ATAVNQECW	0	FQMDYSLEY	0	ADPEYFDEPPRPELP	0
ASYAAAAAY	1	DTRAIDQFF	0	GMFANRWII	0	AEWILKTLVNTEHAY	0
ASYAGAGAY	1	ESLLHQASW	1	GMFSWNLAY	0	AFLICLIKVLRGQIV	1
ASYGAGAGY	1	ESPSSDEDY	0	HMMAVTLFY	0	AHGRKRIVCRERLFS	0
ASYQFQLPY	1	FAHDDRYLY	0	HQYPANLFY	0	AHGVFNPEFGPAALS	0
ATAAATEAY	1	FGMPNPEGY	0	ITMVNSLTY	0	ALKLMEKREYKFTCQ	0
ATAWRTGGY	1	FQMDYSLEY	0	KARARLLSM	1	ALLALYAAAIAAAPS	0
ATDFKFAMY	1	FSSQLGLFY	0	KMFHGGLRY	0	ANHCSDAMNIMFEEV	0
ATEDPSSGY	1	FSVPLDEGF	0	KNNFWFWEY	1	APASPEAGAVSTPPV	0
ATIMPHNLY	0	FTFWTFANY	1	KRIRLKHIF	1	APGLPWALQGKRRGA	0
ATTFARFLY	1	FTIRDVLAY	1	KRMMMNLNY	1	AQMHSNNGPQIGSAV	0
ATVKGMQSY	1	HMMAVTLFY	0	KSFFSRLNW	1	AQYRNVWDVDYSAFD	0
ATYQRTRAL	0	HQYPANLFY	0	KSYEHQTPF	1	ARAMLALLALYAAAI	0
AVDVDDGHF	1	HSNLNDATY	0	KTLKGGWFF	1	ASLAHADALAPVVVE	0
AVEDFLAFF	1	HTAEIQQFF	0	LNIMNKLNI	1	ATLSKNKNCILCTVC	1
AVEGGLYPV	1	HTSALSLGY	1	LNWFEIWIV	0	AVHSGPDEYRRLFEP	0
AVSFRNLAY	1	IIYYQLAGY	1	LSNFMLWQF	0	AVSFGCAVFPRAGET	0
CSDETTLYY	1	ISRQIHWCW	0	MARWITWAM	1	CAVFPRAGETFEVRF	0
CTDDNALAY	1	ISVQPLWEW	0	MMHASTSPF	0	CDGLCVPPEARLAWS	0
CTELKLSDY	1	ITLKVFAGY	1	MQYLNPPPY	0	CRYDKDADINVVTQA	1
CTLNKSHLY	0	KSAAIDGEY	0	RARKRGITM	1	DAADALAPSLRCEAV	0
DSDDWLNKY	0	KSLDNYQEW	0	RAYRNALSM	1	DAETEGVYTWRVLSA	0
DTEDNVPPW	1	KSWPAAIDW	0	RIRAANLPI	1	DAMNIMFEEVFNTDF	0
EIAQHGAWY	1	MANIFRGSY	0	RIYSHIAPY	0	DGLDAMEPDTAPGLP	0
EISGSSARY	1	MLYPRVWPY	0	RLASYGLYY	0	DHAADTVYHLGACAE	0
ESDMEVFDY	1	MNYAAAAAY	1	RLFFIDWEY	0	DKGFVLGHSITDVTF	0
ESENISEPY	0	MTAASYARY	1	RLRRRRHPL	0	DTTPNGGGGGNSSEG	0
ESSDDELPY	0	MTAHITVPY	1	RLYPFGSYY	0	DYRYAISSTNEIGLL	1
ETESVNSNY	1	MTRGLLGSY	1	RMFKRVFNM	1	DYTCRLEGLPSQLPV	0
EVAGAGSGF	1	MTRVTNNVY	0	RMFLAMITY	0	DYYPRRSVRLRWFAD	0
EVDQTKIQY	1	MVASQLARY	1	RNMSRIFPY	1	EAGAVSTPPVPPPSV	0
FLYPSWSLY	1	MVFQNYALY	1	RNNDPTLPY	0	EALKPHFKSLGQTIT	0
FSIPVTFSY	0	MYADDTAGW	0	RQHPGLFPF	0	EAMTNYAKEGIQFMK	1
FTAMQALDY	1	QQYHRFGLY	1	RRARYWLTY	1	EGFAVCDGLCVPPEA	1
FTDNNELEF	1	QTWHGDAPY	0	RRFFPYYVY	1	EHPVDAAFVTNSTVA	0
FTFWTFANY	1	QTYMYTGQY	0	RRFKYLLNV	1	EIRPMEKVRAGKTRI	0
FTIRDVLAY	1	RSVWIPGRW	1	RRFNRTKPM	0	EKREYKFTCQTFLKD	0
FTYAPAGMY	1	RVYNNTARY	0	RRLHRLLLM	1	EKVRAGKTRIVDVLP	0
GSDGGLDDY	1	RVYPNPEVY	0	RRMATTFTF	1	ELSPSPPPTPAPASP	0
GSQYVSLAY	0	RYQAQQVEW	0	RRRQWASCM	1	ENALLVALFGYVGYQ	1
GTDSGFAAY	1	SAYYLDIGF	0	RRSRRSLTV	1	ENKRITVEGGMPSGC	0
GTDSNGMLW	1	SLRPNDIVY	0	RRVRRRVLV	0	EPFRAVCVVRDYYPR	0
GTEKLTITY	1	SSLPSYAAY	1	RSFRIHILF	0	EPIQLAYNSYETQVP	1
GTEYRLTLY	0	SSMNSDAAY	0	RTFDRFFEE	0	ERLFSARVGDAVSFG	0
GTFDLGGLY	0	SSNAKNSEW	0	RTLDTLALY	0	ERVHVMRKTKLAPTV	0
GTTEVNGLY	0	SSNPVMSRF	0	RVFKETLFL	0	FDEPPRPELPRERLL	0
GVEPGHAFY	1	SSVGVTHGY	1	RVFNNYMPY	1	FEDTQRYDASPASVS	0
HIASMRRNY	1	SSVSSFERF	0	RVFYFAIFY	0	FEVRFYRRGRFRSPD	0

HMMAVTLFY	0
HSDDALFIY	1
HTAAPWGSY	0
HTSALSLGY	1
IAGFIEGGW	0
ISAYTHWYY	1
ITAGYNRYY	0
ITDITKYLY	1
ITDITSPLW	1
ITDYIVGYY	1
ITFQSILGY	0
ITTFFTFAY	0
IVDCLTEMY	1
IVDINVKDY	1
KIAPLMVAY	1
KLDAWLLPF	1
KLDPTNTLW	1
KMARLGKGY	1
KMFHGGLRY	1
KSDGTGTIY	1
KSDLQPPNY	1
KSNRIPFLY	1
KSTDSESDW	0
KTAVVVTRY	1
KTFEWGVFY	1
KVFFGPIYY	1
LIDGRTSFY	1
LLDGLLAWY	1
LSDDAVVCY	1
LSTASSWSY	1
LTAHYCFLY	0
LTDDMIAAY	1
LTDSDSPTY	1
LTFLDCLYY	0
LTMDREMLY	0
MADSFKSDY	1
MIDSDEWVY	1
MIEPRTLQY	1
MIGGIGRFY	1
MLASIDLKY	0
MSAIVSCRY	1
MSNEGSYFF	0
MSSAAHLLY	1
MSWESTAEY	1
MTAASYARY	1
MTAHITVPY	1
MTRGILGSY	0
MTRGLLGSY	0
MTRVLPFTY	0
MTRVTNNVY	0
MTSGSSSGF	1
NADTLCIGY	1
NIDNMCHLY	1
NSDTVGWSW	1
NTDAFSREY	1
QIGNIISIW	1
QSAANMYIY	0
QTDNDIWFW	1
QTDNQLAVF	1
QTDPLWQKY	1
QTEENLLDF	1
QTNLYNLLY	1
STEPPMLNY	0
STFATVLEY	0
TVYNGTSKY	0
VSIPVTNTW	1
VSRLEHQMW	0
VSYAAAAAY	1
VTEPGTAQY	0
YANMWSLMY	0
YAQMWSLMY	0
YAYNSSLLY	0
YLSGIAQYY	1
YSRMLYIEF	0
YSYIFLSSY	0
YTASVVAAY	1
YTGPDHQEW	0
YTIGIGAFY	1
YTITYHDDW	0
YTNPQFNVY	0
YTSDYFISY	1
RVRRLNWAA	1
RVYPNPEVY	0
RWFVRNPFF	0
RYFTVAFLF	0
SARRRHLVF	1
SMFDSWGPF	0
SQYHRFPIY	0
SRWSRKMLM	0
SSMNSFLLY	0
TSFASSWIY	1
TTRHRKPTY	0
TVFYNIPPM	0
VMFRNASEY	0
VSYAAAAAY	0
VTFWGFWLF	0
YAYNSSLLY	0
YMIGYTAYY	0
YRFRFRSVY	1
YSRPWNWTF	1
YSYIFLSSY	0
YTFFFTQYF	0
YTYATRGIY	0
FGTHFAQYRNVWDVD	0
FNTDFGFHPNAEWIL	0
FQGLFEIPSYRSLYL	0
FRSPDADPEYFDEPP	0
FSEALRPHVYHPAAV	0
FSSANASLAHADALA	0
GACAEHPGLLNVRSA	0
GCNPDVDWQRFGTHF	0
GDKATAHGRKRIVCR	0
GERATVANVSGEVSV	0
GEVSVRVAAADAETE	0
GFHPNAEWILKTLVN	0
GGGGGNSSEGELSPS	0
GKTRIVDVLPVEHIL	0
GLAAADAADALAPSL	0
GLIVDTRDVEERVHV	0
GLSAPPVLFGEPFRA	0
GLSTAENALLVALFG	1
GQTITPADKSDKGFV	0
GVYTWRVLSANGTEV	0
HFKSLGQTITPADKS	0
HMDYGTGFYKPVMAS	0
HPAAVSVRFVEGFAV	0
HPGLLNVRSARPLSD	0
HSVLGTANAPLSTYE	0
ICAAGSFKVTALNVV	1
IFSKHKGNTKMSEED	0
IGRFCAQMHSNNGPQ	0
IGSAVGCNPDVDWQR	0
ILAIVLVIMATCVYY	0
ILSFARRGTIQEKLI	0
INTILNNIYVLYALR	0
KAILISCISNKWQFI	1
KALFRRCAADYASRL	0
KFTCQTFLKDEIRPM	0
KGNTKMSEEDKALFR	0
KHKVRNEVMVHWFDD	1
KKFFLLSSRVKELII	1
KKGKNFSFAGTIIEG	1
KLASSAFSGLFG	1
KTLEAILSFARRGTI	1
KTLVNTEHAYENKRI	0
LAPSLRCEAVWYRDS	0
LAPTVAHGVFNPEFG	0
LDGPVDYTCRLEGLP	0
LEGLPSQLPVFEDTQ	0
LGHSITDVTFLKRHF	0
LIKVLRGQIVQGVIW	1
LNEGVVLDEVIFSKH	0
LVIMATCVYYRQAGP	0
LYALRRHYEGVELDS	0
MEPDTAPGLPWALQG	0
MFEEVFNTDFGFHPN	0
MMASLARAMLALLAL	0
MPSGCSATSIINTIL	0
MRKTKLAPTVAHGVF	0
MSEEDKALFRRCAAD	0
NDWFSKLASSAF	1
NGTEVRSANVSLLLY	0
NGTVGPEVEAALKLM	0
NKDPRLNEGVVLDEV	0
NNGPQIGSAVGCNPD	0

QTWHGDAPY	1
QVSRPMFLY	1
RADSMMLGY	1
RIARFHRPY	1
RLASYGLYY	1
RMFLAMITY	0
RSADGSPPY	1
RTDAWSYPV	1
RTLASGLIY	1
RTWAYHGSY	1
RTWFHGSLY	1
RTWHYCGSY	0
RTWNYHGSY	1
RVERIKSEY	1
RVFPGDHFY	1
RVSTGLYRY	1
SSDDIPPRW	1
SSDISFIKY	1
SSDLRSWTF	1
SSFERFEIF	0
SSMNSFLLY	1
SSSFSFGGF	1
SSVGVTHGY	1
STAPTGSWF	1
STFATVLEY	1
STYQPLPLY	1
SVAMCRTPF	1
SVDGFRASY	1
SVEMNAPNY	1
SVEVKLPDY	1
TIDKSSPLY	1
TLELRSRYW	1
TMDVNHPIY	1
TQDLFLPFY	1
TSDGFINGW	1
TSSARSSEW	0
TTSDFFVNY	0
TVYNGTSKY	1
VAGGTGSVY	1
VLDKWNTNY	1
VSALRLFNY	0
VSDGGPNLY	1
VSDGPPTGY	1
VSFNQNLEY	0
VSYAAAAAY	0
VTDPGGLYY	1
VTIGNAYIY	0
VTRGAVLMY	0
VVAANRSAF	1
VVDALRNIY	1
WSQDPTMLY	0
WTGMVDGWY	0
WVAGVQLLY	0
YAQMWTLMY	1
YIFFASFYY	0
YLSGIAQYY	1
YSAEALLPY	1
YSYIFLSSY	0
YTASVVAAY	1
YTDKIAMSY	1
YTFEPHYFY	0
YTSDYFISY	1

NNIYVLYALRRHYEG	0
NPEFGPAALSNKDPR	0
NSSEGELSPSPPPTP	0
NSTVADELGRRTRVS	0
NVRSARPLSDLDGPV	0
PAALSNKDPRLNEGV	0
PADKSDKGFVLGHSI	0
PASVSWPVVSSMIVV	0
PDEYRRLFEPFQGLF	0
PEVEAALKLMEKREY	0
PPPSVSRRKPPRNNN	0
PPPTPAPASPEAGAV	0
PRNNNRTRVHGDKAT	0
PVLFGEPFRAVCVVR	1
PVMASKTLEAILSFA	0
PVVVEGERATVANVS	0
QEKLISVAGLAVHSG	0
RADVPGLAAADAADA	0
RAGETFEVRFYRRGR	0
RCAADYASRLHSVLG	0
RCEAVWYRDSVASQR	0
RERLLFSSANASLAH	0
RHYEGVELDSYTMIS	0
RIVCRERLFSARVGD	0
RLAWSDHAADTVYHL	0
RLFEPFQGLFEIPSY	0
RPELPRERLLFSSAN	0
RPHVYHPAAVSVRFV	0
RPLSDLDGPVDYTCR	0
RQLSSNYILELLYKF	1
RRGTIQEKLISVAGL	0
RSANVSLLLYSQPEF	0
RSVRLRWFADEHPVD	0
RTRVHGDKATAHGRK	0
RTRVSVVNVTRADVP	0
RVAAADAETEGVYTW	0
RVLSANGTEVRSANV	0
RWFADEHPVDAAFVT	0
RYDASPASVSWPVVS	0
SATSIINTILNNIYV	0
SEGKIRQLSSNYILE	1
SLLLYSQPEFGLSAP	0
SMIVVIAGIGILAIV	0
SQLPVFEDTQRYDAS	0
SQPEFGLSAPPVLFG	0
SQSPYVVVATNAIES	1
SRRKPPRNNNRTRVH	0
STPPVPPPSVSRRKP	0
SVAGLAVHSGPDEYR	0
SVRFVEGFAVCDGLC	1
SYFQQYMLKGEYQYW	1
TANAPLSTYEAIKGV	0
TASALYLISYYVIPQ	1
TDVTFLKRHFHMDYG	0
TEHAYENKRITVEGG	0
TFLKDEIRPMEKVRA	0
TGFYKPVMASKTLEA	0
TLILILPTYELTKLY	1
TRDVEERVHVMRKTK	0
TRVWNSASTTAFLIC	1
TTLLNGSAFYLVCPI	1
TTTALDTTPNGGGGG	0

YTYPCIPEY	1	TVEGGMPSGCSATSI	0
YVFVGSSRY	0	TVYHLGACAEHPGLL	0
YVFVGTSRY	1	VALFGYVGYQALSKR	1
YVYFYDLSY	1	VANVSGEVSVRVAAA	0
		VASQRFSEALRPHVY	0
		VCVVRDYYPRRSVRL	0
		VDVLPVEHILYTRMM	0
		VDWQRFGTHFAQYRN	0
		VEHILYTRMMIGRFC	0
		VELDSYTMISYGDDI	0
		VELYYKGTTIKLDFN	1
		VLDEVIFSKHKGNTK	0
		VPPEARLAWSDHAAD	0
		VVASDYDLDFEALKP	0
		VVCQSNNKMTDESEY	1
		VVNVTRADVPGLAAA	0
		VWDVDYSAFDANHCS	0
		WALQGKRRGALIDFE	0
		WPVVSSMIVVIAGIG	0
		WYRDSVASQRFSEAL	0
		YAAAIAAAPSTTTAL	0
		YASRLHSVLGTANAP	0
		YDLDFEALKPHFKSL	1
		YEPRDSYFQQYMLKG	1
		YGDDIVVASDYDLDF	0
		YRRGRFRSPDADPEY	0
		YSAFDANHCSDAMNI	0
		YTMISYGDDIVVASD	0
		YTRMMIGRFCAQMHS	0

Supplemental Table 2. HLA crystal structures.

Allele	PDB
HLA-A*0101	4NQV
HLA-A*0201	3MRE
HLA-A*0201	3MRG
HLA-A*0301	3RL1
HLA-A*1101	1X7Q
HLA-A*2402	1P7Q
HLA-A*6801	4HWZ
HLA-B*0702	4U1H
HLA-B*0801	4QRS
HLA-B*2705	2A83
HLA-B*3501	2CIK
HLA-B*4403	1N2R
HLA-B*5101	1E27
HLA-DRB1*0101	1T5W
HLA-DRB1*0301	1A6A
HLA-DRB1*0401	1J8H
HLA-DRB1*1501	1YYM

Supplemental Table 3. Contact residues in the SLA class II binding pockets based on HLA contacts (Hc).

SLA position	Pocket (peptide position) A(1)	B(4)	C(6)	D(7)	E(9)
9H					■
11L			■		
13F		■	■		
26L		■			
28E		■	■	■	
30Q			■	■	■
37F					■
38L					■
47Y				■	
57D					■
60D					■
61W				■	■
67L			■	■	
70Q		■		■	
71R		■	■	■	
74E		■	■	■	
78Y		■			
81H	■				
85I	■				
86L	■				
89F	■				
90L	■				

Positions of the residues in the SLA binding pockets are shown. The first column (SLA position) is the residue and position in the SLA-DRB1*0201 protein sequence (Genbank:61652983). The next columns show, shaded in gray, the positions involved in pockets A through E that interact with specific positions of the peptides (peptide position).

Supplemental Table 1. GenBank identification numbers of gene sequences of proteins expressed by representative swine IAV.

	GenBank id							
Strain	**PB2**	**PB1, PB1-F2**	**PA**	**HA**	**NP**	**NA**	**M2, M1**	**NS2, NS1**
A/California/04/ 2009(H1N1)	227809823	332384205	227977113	229535948	332384201	332384199	332384196	332384209
A/swine/Illinois/5265/ 2010(H1N1)				290873719		290873721	290873723	
A/swine/Ohio/511445/ 2007(H1N1)	197344175	197344177	197344180	197344171	197344182	197344173	197344184	197344187
A/swine/Minnesota/ 02011/2008(H1N2)	304272397	304272373	304272351	304272327	304272303	304272279	304272249	304272213
A/swine/Minnesota/ A01301731/2012(H1N2)	421920014	421920016	421920018	421920020	421920022	421920024	421920026	421920029
A/swine/Texas/4199-2/ 1998(H3N2)	340784749	340784753	340784756	340784760	340784762	340784766	340784768	340784771
A/turkey/Ohio/313053/ 2004(H3N2)	361630452	367462658	367462657	371574609	367462656	370321100	361630450	361630451

Supplemental Text 1. Concatemer construct sequences.

>Class I concatemer construct
cggatctaccagattctggccatctacatcaccatcggcaagtgccccaagtactctctgctgaccgaagtggaaacatatggccccggacccggcac
cagtgccgatcagcagtctctgtacggacccggacccggggagatgaacgccccccaactatcactatctggccagctgcatgggactgatctacgac
accgtgaaccggacccaccagtacggcaccgagaagctgaccatcacatacgccttcgacgagcggcggaacaaatacgataccgtgcacgacc
ggacccccctatctgactgaagtggaaacttacgtgctgggacccggacccggcagcgtgaagaatggcacatacgactacggcaccatcaaggac
agaagcccctacaacgccgacacactgtgcatcggctatgggcccggacccgggggaatggtggatgggtggtacggctactgcaccgagctgaa
gctgagcgactatggacccggacccggcgtgtcagatggcggcccaaatctgtacgaaatctgccccaagctggccgagtacaagagctgcatcaa
ccggtgcttctaccaagtgtcccggcccatgtttctgtacggcgccaaagaagtggctctgagctacggacccggacccggcgaactggatgcccca
aactaccactacgatctgctggaaaatctgcaagcctacgccagccaaggcaccaagcggagctacaacaccgatctggaagctctgatggaaaac
atggacaaggccgtgaagctgtatctgagcaccgccagcagctggtcttacggacccggacccggcggaatgatcgacgggtggtatggatactga
tga

>Class II concatemer construct
acaagaggcgtgcagatcgccagcaacgagaacgtggaaaccatggacagcaacacactggaactgcggacatacgtgctgagcatcatcccca
gcggccctctgaaggccgagatcgcccagagactggaaagcgtgtacagatacggcttcgtggccaacttcagcatggaactgcccagcttcggcg
tgtccggctgcagaacattctttctgacacaaggcgctctgctgaacgacaagcacatcgacccccttcaagctgctgcagaatagccaagtggtgtctc
tgatgcggccctacgaggaactgcgcgagcagctgtccagcgtgtccagcttcgagcggttccaagacattctgatgcgcatgagcaagatgcagct
gggctccagcagcgagatgatgggcatgttcaatatgctgagcaccgtgctgggcgtgtccatcggcgacaagatcacattcgaggccaccggcaa
tctggtggtgccccggtacagatccaagtttctgctgatggacgctctgaagctgtccatcgaggaccctagctgcatgggactgatctacaaccggat
gggcaccgtgaccaccgaggccgcttttggactcgtgtgcttcgagcagatcactttcatgcaagctctgcagctgctgctggaagtggaaaaccaga
cttacgtgaacatcagcaacaccaacttcgccgctggccagagcgtggtgtccgtgaagctggtgcccagatacgcctttgccatggaacggaatgc
cggcagcggcatcatcatcagcgaagtgcacatctactatctggaaaaggccaacaagatcaagagcgagaaaacccacatccacatctttggcccc
ggacccggcatggccaatctgattctgcagatcggcaacatcatctccatctggatcagccactctatcgaggatctgatctttctggctcggagcgcac
tgattctgagaggcagcgtggcccacaagagctgtctgcccaccagacagatggtgcacgccatgagaaccatcggcacccaccctagctctagcg
cctccgtggtgtctgtgaaactggccggcaatagctctctgtgccccgtgtccacccggatctaccagattctggccatctacagcaccgtggcctcttc
tctggtgctcgtgtgatga

Supplemental Table 2. Low-resolution SLA-typing results.

		SLA class I			SLA class II		
Group	**Pig**	**SLA-1**	**SLA-2**	**SLA-3**	**DRB1**	**DQB1**	**DQA**
NV	413	08XX	0901-02	05XX,0602	07XX,10XX	0201,06XX	01XX,02XX
	414	08XX,15XX	12XX,16XX	0601,07XX	0401-02,0403-04	0202,0302-03	02XX
Sham	419	08XX,15XX	05XX,16XX	07XX	0403-04,10XX	06XX,07XX	01XX
	420	08XX	0901-02,12XX	05XX,0601	0401-02,07XX	0201,0202	02XX
	421	08XX,12XX,1301	05XX,10XX	05XX,07XX	06XX,10XX	06XX,07XX	01XX
	422	08XX,15XX	12XX,16XX	0601,07XX	0401-02,0403-04	0202,0302-03	02XX
	423	08XX	0901-02,12XX	05XX,0601	0401-02,07XX	0201,0202	02XX
	424	08XX	12XX	0601	0401-02	0202	02XX
	425	12XX,1301	10XX	05XX	06XX,10XX	06XX,07XX	01XX
	426	08XX	05XX,12XX	0601,07XX	0401-02,10XX	0202,06XX	01XX,02XX
PigMatrix-	428	08XX	05XX,12XX	0601,07XX	0401-02,10XX	0202,06XX	01XX,02XX
EDV	429	08XX,12XX,1301	0901-02,12XX	05XX,0602	06XX,10XX	06XX,07XX	01XX
	430	08XX	05XX,12XX	0601,07XX	0401-02,10XX	0202,06XX	01XX,02XX
	431	08XX,12XX,1301	05XX,10XX	05XX,07XX	06XX,10XX	06XX,07XX	01XX
	432	08XX	05XX,10XX	05XX,07XX	10XX	06XX	01XX
	433	08XX,12XX,1301	0901-02,12XX	05XX,0602	06XX,10XX	06XX,07XX	01XX
	434	04XX,08XX	04XX,12XX	04XX,0601	0401-02,09XX	0202,08XX	02XX,03XX
FluSure	435	1103,12XX,1301	10XX,jh02	05XX	06XX	07XX	01XX
	436	08XX	05XX,12XX	05XX,0601	0401-02,10XX	0202,06XX	01XX,02XX
	437	12XX,1301	10XX	05XX	06XX	07XX	01XX
	438	08XX,12XX,1301	05XX,12XX	05XX,0601	0401-02,06XX	0202,07XX	01XX,02XX
	439	08XX	12XX	0601	0401-02	0202	02XX
	440	07XX,08XX	02XX,05XX	04XX,07XX	02XX,10XX	0201,06XX	01XX,02XX
	441	1103,12XX,1301	10XX,jh02	05XX	06XX	07XX	01XX

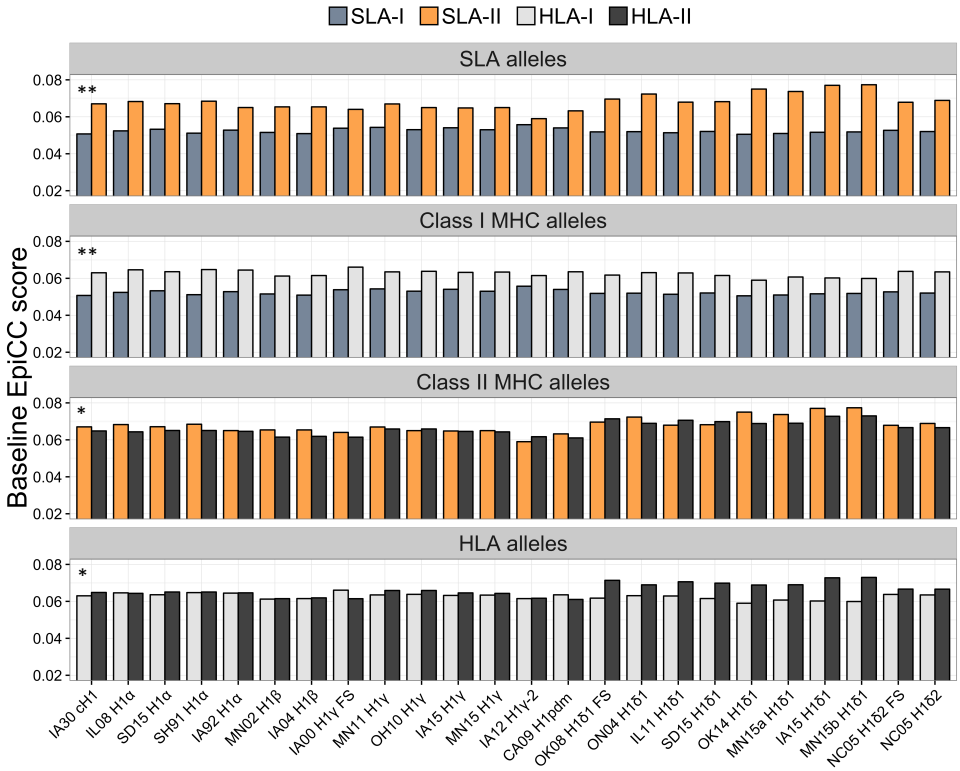
Supplemental information for Chapter 4

Supplemental Table 1. EpiCC scores between HA sequences of IA00 H1γ FS vaccine virus and field viruses.

Strain[a]	Score			
	Shared (%)[b]	Vaccine unique	Strain unique	Total EpiCC
IA00 H1γ FS	0.118 (100)	0.000	0.000	0.118
MN11 H1γ (P)	0.097 (82.2)	0.022	0.025	0.050
IA92 H1α (P)	0.092 (78.0)	0.027	0.027	0.038
OH10 H1γ (P)	0.091 (77.1)	0.028	0.028	0.036
IA15 H1γ	0.088 (74.6)	0.030	0.031	0.026
CA09 H1pdm (P)	0.087 (73.7)	0.032	0.031	0.023
MN15 H1γ	0.085 (72.0)	0.034	0.033	0.018
SH91 H1α	0.083 (70.3)	0.035	0.037	0.011
IA04 H1β	0.079 (66.9)	0.040	0.039	0.000
MN02 H1β (P)*	0.078 (66.1)	0.041	0.040	-0.002
IA12 H1γ-2	0.071 (60.2)	0.048	0.045	-0.023
IL08 H1α (PP)*	0.068 (57.6)	0.051	0.054	-0.038
IA30 cH1	0.060 (50.8)	0.060	0.060	-0.060
SD15 H1α	0.058 (49.2)	0.062	0.065	-0.068
NC05 H1δ2 FS	0.053 (44.9)	0.066	0.069	-0.082
NC05 H1δ2	0.053 (44.9)	0.066	0.069	-0.083
OK08 H1δ1 FS	0.053 (44.9)	0.066	0.070	-0.083
ON04 H1δ1	0.052 (44.1)	0.068	0.074	-0.090
IL11 H1δ1	0.050 (42.4)	0.070	0.071	-0.091
OK14 H1δ1	0.052 (44.1)	0.068	0.076	-0.092
SD15 H1δ1	0.049 (41.5)	0.071	0.073	-0.095
IA15 H1δ1	0.049 (41.5)	0.071	0.081	-0.103
MN15b H1δ1	0.048 (40.7)	0.072	0.083	-0.107
MN15a H1δ1	0.043 (36.4)	0.077	0.083	-0.117

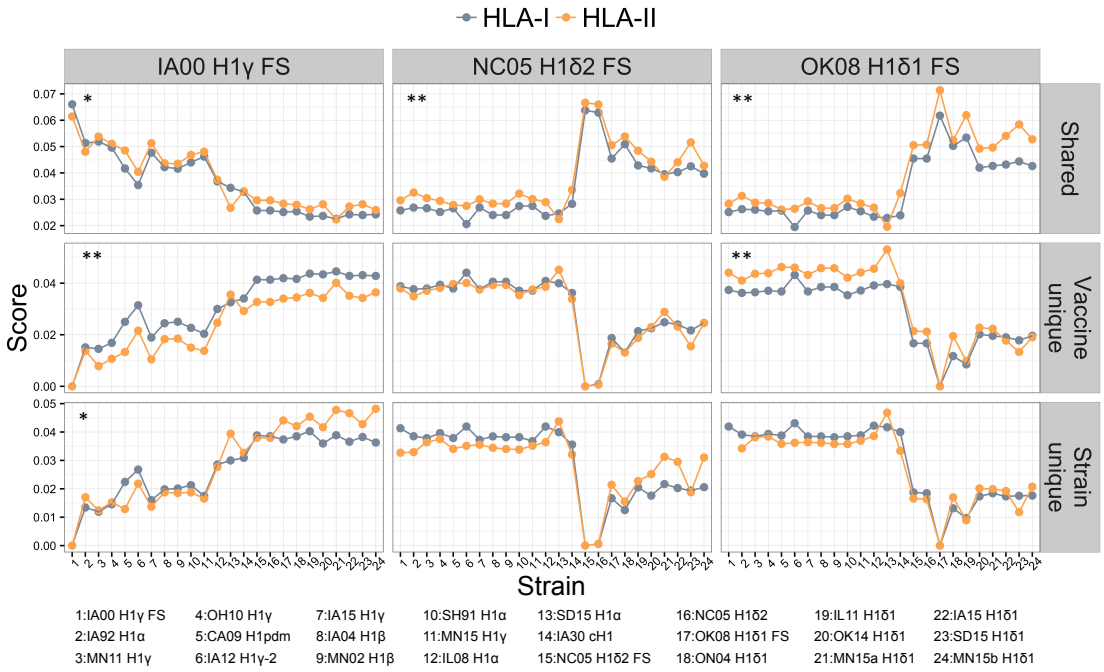
[a]FS γ-cluster vaccine strain was protective or partially protective against challenge with viruses annotated as (P) or (PP), respectively. Viruses used to set the thresholds are marked with an asterisk (*).
[b]Ratio of the score of shared epitopes relative to the baseline EpiCC scores expressed as percentage.
Table is sorted by total EpiCC score. Strains below the total EpiCC score threshold associated with partial protection are shown in gray.

Supplemental Fig. 1. Comparison of HA baseline EpiCC score by set of MHC alleles.



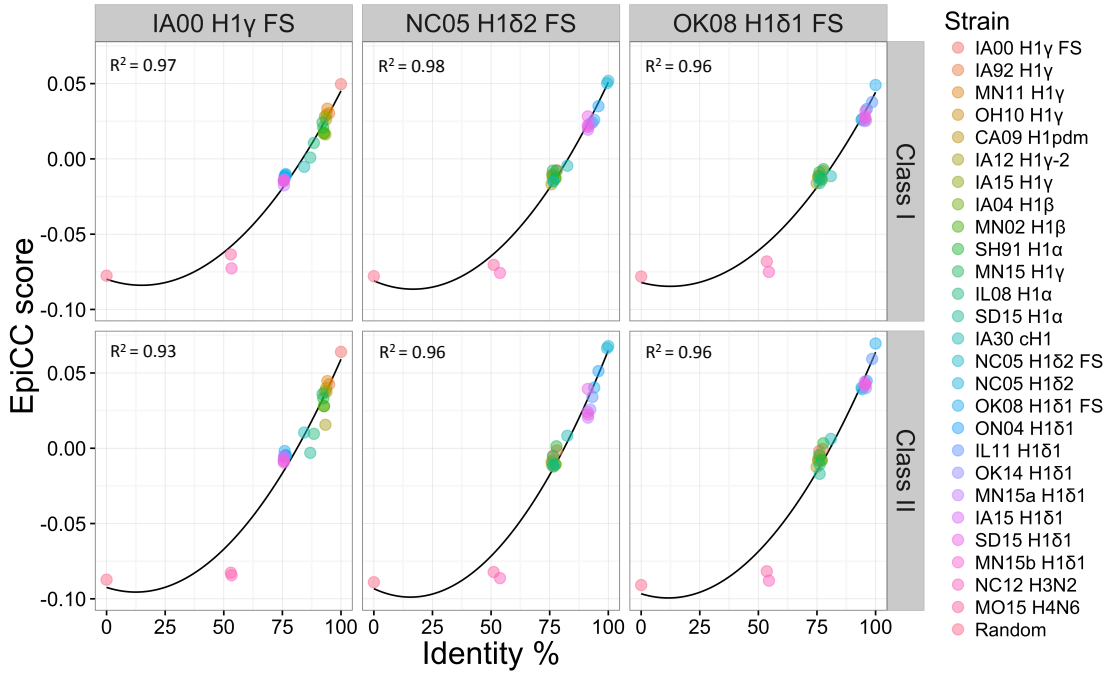
Baseline EpiCC score of HA sequences of each virus ($E(s,s)_A$) was determined using predictions for swine and human class I and II MHC alleles (SLA-I: 8 alleles, SLA-II: 5 alleles, HLA-I: 6 alleles, and HLA-II: 8 alleles). P-values of comparisons were calculated using one-tailed Wilcoxon matched-pairs signed rank test (**p<0.001, *p<0.05).

Supplemental Fig. 2. Comparison of scores of shared and unique epitopes across strains.



Scores of shared, vaccine unique and strain unique HLA class I and II epitopes were determined for the comparison of HA sequences from vaccine viruses and field (challenge) strains. Note that y-axes show different scales. Solid connecting lines are included only for visualization purposes. P-values of comparisons were calculated using one-tailed Wilcoxon matched-pairs signed rank test (**p<0.001, *p<0.05). HA vaccine sequences had higher scores for shared epitopes with strains belonging to the same H1 cluster or the same HA lineage. Scores of class II shared epitopes were significantly higher than those of class I. Class II scores of unique epitopes were also higher for OK08 H1δ1 FS (vaccine unique) and NC05 H1δ2 FS (strain unique). In contrast, class I scores of IA00 H1γ FS vaccine unique epitopes were higher than those of class II, which is explained by the higher baseline EpiCC score of the vaccine. Viruses are sorted by nucleotide identity relative to H1γ FS. Strain numbers on the x-axis are described in detail in the legend below.

Supplemental Fig. 3. Relationship between EpiCC scores and nucleotide identity.



The second order polynomial relationship between class I (top), II (bottom) EpiCC scores, and nucleotide identity for each FS virus are shown. R^2 of regression models are shown. H3N2, H4N6 and a random sequence were included in this analysis to represent the lower end of the identity range. Class I and II EpiCC scores correlated with identity between HA nucleotide sequences of vaccines and viral strains (r=0.89 – 0.90). However, their relationship was non linear; instead, it was second order polynomial (R^2=0.93 – 0.98). EpiCC score was close to the lowest at approximately 50% identity.